JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 17, No. 5, September–October 1994

Maximum a Posteriori Image Registration/Motion Estimation

Yaakov Oshman* and Baruch Menis†
Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel

A novel method is presented for on-line, recursive image registration/motion estimation. Comprising a Kalman filter and an interlaced new image registration algorithm, the method enables the estimation of aircraft motion from a sequence of terrain images, acquired by an airborne, down-looking electro-optical sensor. Contrary to other methods of vehicle motion estimation that are based on the conventional mean-of-squared-differences (MSD) image registration algorithm, the new method utilizes the maximum a posteriori (MAP) estimation methodology to draw statistical information from the prediction level of the Kalman filter. The resulting image registration algorithm is thus rendered more accurate and robust with respect to loss of lock (measurement divergence). A small error statistical analysis shows that, under reasonable conditions, the new MAP algorithm is unbiased and efficient, and its estimation error covariance is smaller than that of the ordinary MSD algorithm. The superiority of the new MAP algorithm over the conventional MSD algorithm was substantiated via an extensive experimental investigation, using real aerial photographs.

I. Introduction

THIS paper is concerned with the problem of on-line and recursive estimation of vehicle motion from a sequence of terrain images, acquired by an airborne, down-looking electro-optical sensor. Deriving motion information from visual data is a well-known and important problem in many applications, e.g., in the design of completely passive and autonomous navigation systems[1,2] and in computer vision and robotics.[3] Essentially, this estimation task is performed by solving in a recursive manner the problem of image registration,[4] defined as the following: given a sequence of image brightness arrays, acquired by an onboard sensor, estimate the relative shift between each pair of consecutive images in the sequence. Since the interframe shift acquired at a fixed time interval is linearly related to the ratio of ground velocity to ground altitude (v/h), this ratio can be estimated by recursively solving the image registration problem for each pair of frames in the acquired image sequence of the changing terrain. Moreover, by aiding this estimate with additional measurements, obtained from other onboard sensors, the ground speed can be estimated and used in an inertial navigation system.

Most methods, which use visual information to estimate vehicle motion, compute an estimate of the two-dimensional field of instantaneous velocities of brightness values (gray levels) in the image plane, defined as the optical flow. When used in conjunction with added constraints or information regarding the scene and/or camera motion, the estimated optical flow yields an estimate of the actual three-dimensional relative motion between the scenery and the sensor. Methods to estimate the optical flow field lie within two general classes. Gradient-based methods[5] compute an estimate of the optical flow field over the entire image, utilizing a relationship between the motion of surfaces and the derivatives of image brightness. Feature-based approaches[6] locate and track over time a set of highly discriminatory, two-dimensional features (regions) in the images corresponding to three-dimensional objects in the scene, such as corners, occluding boundaries of surfaces, etc. The method presented in this paper belongs to the category of feature-based methods.

Optical flow computation is, in general, expensive and in most cases rather noisy. To alleviate the computational burden problem, image registration methods have been developed[4] that utilize only a small portion of the image, so that only a local estimate of the optical flow is computed using some subset of the image pixels. To overcome the noisy character of the optical flow estimate, attention has been recently focused on filtering the output of the optical flow estimator by utilizing this estimator as a generalized sensor within a recursive Kalman filtering framework.[3,7,8] The Kalman filter performs an optimal fusion of the information derived from the optical flow with prior information regarding the vehicle motion. Since a Kalman filter is used, the resulting vehicle motion estimate is inherently based on information contained in the entire image sequence (measurement history), rather than just on information contained in isolated pairs of images. Moreover, the optical flow measurements are filtered, and the motion estimation error is thus statistically decreased.

This paper presents a new image registration/motion estimation algorithm, that joins a Kalman state estimator with an optical flow processor similar to the mean-of-squared-differences (MSD) shift estimator used in Ref. 2. In that regard it belongs to the class of Kalman filter-based motion estimation algorithms discussed earlier. However, whereas in other members of this class the shift estimator is just used to drive the Kalman filter measurement update, e.g., Ref. 7, in the new scheme the Kalman prediction statistics are fed back to drive a modified image registration algorithm. This modified shift estimator fuses the a priori information provided by the Kalman predictor with the optical flow information to form a maximum a posteriori (MAP) estimate of the interframe shift. This MAP estimate is then utilized, in the conventional manner, by the measurement update level of the Kalman filter, along with measurements from other onboard sensors. Operating on a modified cost function which contains the a priori information contributed by the Kalman predictor, the new MAP algorithm is rendered more robust than the conventional shift estimator, with respect to "loss of lock" and "false lock." [9]

In the next section, the image statistical model is defined, and the interlaced Kalman filter/MSD estimator scheme is discussed. A small error statistical analysis of the MSD shift estimator is presented in Sec. III, followed by a derivation and analysis of the new Kalman filter/MAP estimation scheme. The experimental program performed to demonstrate the performance of the new algorithm is presented in Sec. V. Concluding remarks are offered in the last section.

II. Motion Estimation via MSD Image Registration

In this section the image statistical model is presented. The classical MSD image registration algorithm is discussed along with its usage within the framework of a Kalman motion estimator.

Statistical Image Model

In what follows, the terrain image is assumed to be projected onto the image plane of a stabilized, down-looking electro-optical sensor.

Received June 5, 1993; revision received Oct. 6, 1993; accepted for publication Oct. 13, 1993.

*Senior Lecturer, Head, Flight Control Laboratory, Department of Aerospace Engineering. Member AIAA.

†Graduate Student.

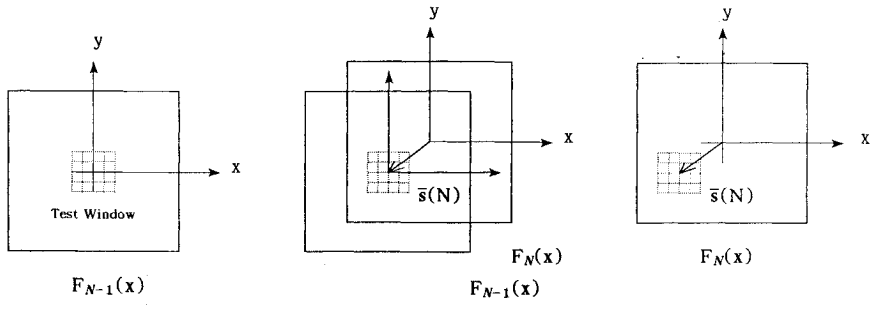


Fig. 1 Image model.

Because of the discrete nature of electro-optical photography, the image space $\mathcal{X}$ is taken as an $L_x \times L_y$ two-dimensional array of digital picture elements (pixels). The location $\boldsymbol{x}$ of each pixel in the image space is determined using an image coordinate system, as shown in Fig. 1. The notation $\boldsymbol{x}_p = (x_p, y_p)^T \in \mathcal{X}$ is adopted. These pixels may each assume one of m brightness (gray) levels. It is assumed that the sensor acquires a sequence of such pictures, at a rate of one picture per T time units. Let $F_N(\boldsymbol{x})$ denote the two-dimensional gray-level array representing the noiseless picture of the scene taken at time NT. $F_N(\boldsymbol{x})$ is assumed to be a discrete representation of a deterministic but unknown, smooth (at least twice differentiable in each direction) two-dimensional surface. Let $\bar{s}(N)$ be the true shift (in units of pixels) between two successive pictures, $F_{N-1}(\boldsymbol{x})$ and $F_N(\boldsymbol{x})$, as shown in Fig. 1. Obviously,

$$F_N[\boldsymbol{x}_p + \bar{s}(N)] = F_{N-1}(\boldsymbol{x}_p) \tag{1}$$

for $\bar{s}(N)$ such that $\boldsymbol{x}_p + \bar{s}(N) \in \mathcal{X}$.

It is assumed that the image is corrupted by an additive noise v, which is due to electronic/thermal camera noise, camera vibrations, and differences between successive pictures due to parallax in nonplanar scenes.[4] Let $v_N(\boldsymbol{x}_p)$ denote the random noise at pixel $\boldsymbol{x}_p$ of the picture taken at time NT. The picture noise is assumed to be a temporally and spatially stationary, Gaussian distributed, zero mean white random field, independent of the pixel brightness level, i.e.,

$$E[v_i(\boldsymbol{x})v_j(\boldsymbol{x}+\boldsymbol{s})] = \begin{cases} \sigma^2 & \text{for } \boldsymbol{s} = 0 \text{ and } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Note that in Ref. 4 the picture noise was assumed to be spatially exponentially correlated. The assumption of a spatially white random field, also used in Ref. 10, corresponds to the more realistic situation of a picture noise having a much wider bandwidth than that of the scene. The assumption that the picture noise is temporally uncorrelated regardless of its source is justified in Ref. 4.

Letting $I_N(\boldsymbol{x})$ denote the measured image (two-dimensional gray level array) taken by the electro-optical camera at time NT, the previous assumptions yield

$$I_N(\boldsymbol{x}_p) = F_N(\boldsymbol{x}_p) + v_N(\boldsymbol{x}_p), \qquad 0 \le I_N(\boldsymbol{x}_p) \le m - 1 \tag{3}$$

Mean-of-Squared-Differences Image Registration Algorithm

The basic image registration algorithm considered in this paper is a pixel-based (iconic) optical flow estimator.[4,7] This algorithm estimates the shift between two successive pictures by searching for a window in the second image, which most resembles (in a mathematical sense to be defined in the sequel) a predefined test window in the first image, as shown in Fig. 1. Using a Fisher approach,[11] the shift $\bar{s}(N)$ between the two consecutive frames $I_{N-1}(\boldsymbol{x})$ and $I_N(\boldsymbol{x})$ is considered an unknown but otherwise deterministic parameter. Let the test window be defined as a subset of n_T pixels $\mathcal{X}_T \subseteq \mathcal{X}$. Also, define an admissible shift as a shift $\boldsymbol{s}$ satisfying

$$\boldsymbol{x}_p \in \mathcal{X}_T \Rightarrow \boldsymbol{x}_p + \boldsymbol{s} \in \mathcal{X}$$

Let $\mathcal{S}$ be the space of all admissible shifts. The MSD algorithm estimates the shift between two successive pictures by performing a two-dimensional search for the minimum of a quadratic "similarity" function, over all admissible shift parameters $\boldsymbol{s} \in \mathcal{S}$, as follows:

$$\hat{s}_{\mathrm{MSD}}(N) = \arg\min_{\boldsymbol{s}\in\mathcal{S}} \frac{1}{n_T} \sum_{\boldsymbol{x}_p \in \mathcal{X}_T} [I_N(\boldsymbol{x}_p + \boldsymbol{s}) - I_{N-1}(\boldsymbol{x}_p)]^2 \tag{4}$$

An efficient, two-step minimization procedure[2] was implemented in this research to evaluate $\hat{s}_{\mathrm{MSD}}(N)$ in an on-line manner. According to this algorithm, the test window $\mathcal{X}_T$ is taken as an $n_x \times n_y$ pixels rectangle, centered about the origin of the picture coordinate system (thus minimizing pixel distortion). For this particular test window, the MSD shift estimate takes the following form:

$$\hat{s}_{\mathrm{MSD}}(N) = \arg\min_{\boldsymbol{s}} \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (I_N[(x_i, y_j)^T + \boldsymbol{s}] - I_{N-1}[(x_i, y_j)^T])^2 \tag{5}$$

The cost function minimization procedure consists of the following two steps.

1) A global search is performed with respect to integer values (in units of pixels) of the shift. A crude estimate of the shift, denoted $\boldsymbol{s}^*$, is found such that the value of the cost (5) is a local minimum in a 3×3 pixel neighborhood of $\boldsymbol{s}^*$. The estimation error associated with $\boldsymbol{s}^*$ is of the order of one pixel. To obtain subpixel accuracy, the next step of the minimization is carried out.

2) A parabolic surface is fitted through the nine cost function values that have been computed in the 3×3 neighborhood of $\boldsymbol{s}^*$. The minimum of this surface is found using a Newton-Raphson method. A second-order surface was experimentally determined to be satisfactory for test windows larger than 8×8 pixels. [2]

Notice that, since only a small subset of the picture elements is used, a "good" initial condition is required to ensure convergence of the global search performed in the first step of the minimization (a divergence of the search might lead to a false lock or loss of lock). The MAP procedure, presented in the sequel, utilizes statistical information provided by the prediction level of the Kalman filter to alleviate this problem.

The conventional interlaced Kalman filter/MSD estimator scheme is presented next.

Shift Estimator Driven Kalman Filters

The conventional image registration/motion estimation scheme has been proposed by several authors.[2,3,7,10] A key feature of this scheme is the usage of the MSD shift estimator as a generalized sensor, whose output is statistically fused with the output of other sensors via a Kalman filter. The temporal variation of the shift has to be defined for that purpose, using a dynamic model which incorporates available information regarding the motion of the platform (e.g., constant velocity motion, etc.). Since, as will be shown later, the optical flow measurements are temporally correlated, the state estimator to be designed has to take this fact into account. It should be noted in this regard that, although the correlation between successive measurements has been previously pointed out,[7] to the best of the authors' knowledge it has been incorporated into the design of the corresponding Kalman filter only by Bar-Shalom et al.,[10] in the context of target tracking via an imaging infrared sensor.

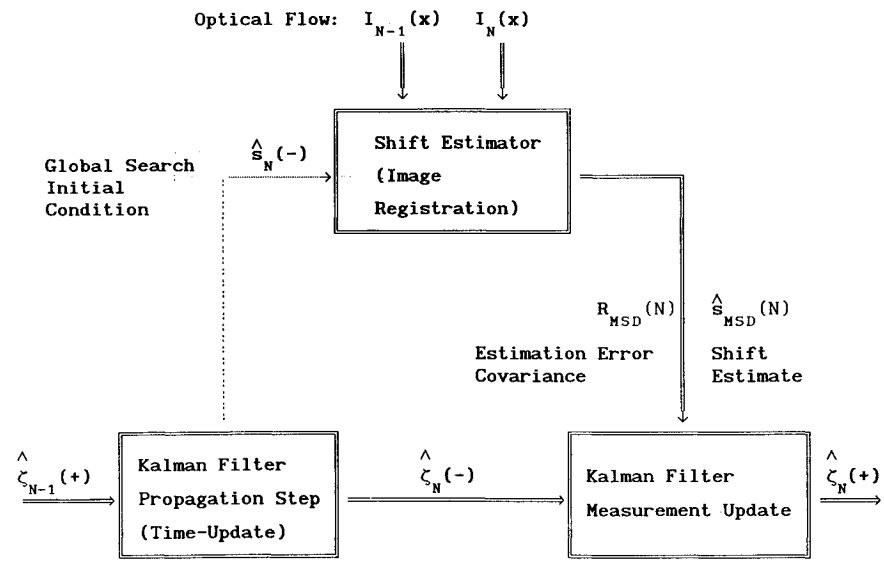


Fig. 2 Motion Estimation via image registration with $\hat{\zeta}(+)$ the a posteriori estimate and $\hat{\zeta}(-)$ and $\hat{s}(-)$ the a priori estimates.

The combined estimation scheme is shown in Fig. 2, where the state of the total dynamic system (comprising the vehicle, payload dynamics, etc.) is denoted ζ. Note that the input from the Kalman prediction step into the shift estimator, namely, the a priori shift estimate, serves only as an initial starting point for the global search of the MSD estimator.

III. Statistical Analysis of the Mean-of-Squared-Differences Shift Estimator

This section presents a small error statistical analysis of the conventional MSD estimator. The main results are summarized in the following theorem.

Theorem 1. Assume the statistical image model presented in the previous section. Assume also that no false lock or loss of lock has occurred (that is, the shift estimator operates in the vicinity of the true shift). Then, the MSD estimator is unbiased and efficient. □

Proof. To assess the statistical efficiency of the MSD estimator, its estimation error covariance has to be compared to the theoretical lower bound given by the Cramer-Rao lower bound (CRLB). Thus, in the first part of the proof, the CRLB is computed, based on the information content in the two successive pictures that are used to estimate the shift. Then, in the second part, unbiasedness is proved and the MSD estimation error covariance is computed and compared to the CRLB.

Explicit computation of the CRLB requires the probability density function of the measurements (conditioned on the shift parameter). To this end, the concept of "measurement" (involved with the process of optimally allocating an area within the second picture that most resembles the test window contained in the first picture) is first defined.

Let the true two-dimensional shift between the two frames under consideration be $\bar{s}(N)$, and let $\mathcal{X}_T$ be a rectangular test window of $n_x \times n_y = K$ pixels (to facilitate the notation, the shift time dependence will be omitted in the sequel unless explicitly needed). To simplify the ensuing development, let $\boldsymbol{x}(i, j)$ denote the location of the pixel whose coordinates, in the picture coordinate system, are (x_i, y_j), i.e., $\boldsymbol{x}(i, j) \triangleq (x_i, y_j)^T$. For an assumed shift $\boldsymbol{s} \equiv (s_x, s_y)^T$, the measurement at the pixel located at $\boldsymbol{x}(i, j)$ is defined as the shifted image difference, i.e.,

$$z_N[\boldsymbol{x}(i, j), \boldsymbol{s}] \triangleq I_N[\boldsymbol{x}(i, j) + \boldsymbol{s}] - I_{N-1}[\boldsymbol{x}(i, j)] \tag{6}$$

Performing K measurements (for the same value of the assumed shift) at K different pixels, we define

$$\boldsymbol{Z}_N^K \triangleq \{z_N[\boldsymbol{x}(i_1, j_1), \boldsymbol{s}], z_N[\boldsymbol{x}(i_2, j_2), \boldsymbol{s}], \ldots, z_N[\boldsymbol{x}(i_K, j_K), \boldsymbol{s}]\}^T \tag{7}$$

as the total measurement vector. With these definitions on hand, the CRLB is computed next. As is well known, the estimation error covariance of any estimator $\hat{\boldsymbol{s}}$ of the true interframe shift $\bar{\boldsymbol{s}}$ is bounded by the CRLB as follows:

$$\operatorname{cov}\{\hat{\boldsymbol{s}} - \bar{\boldsymbol{s}}\} \geq \mathcal{F}^{-1}$$

where $\mathcal{F}$ is the Fisher information matrix (FIM). For an unbiased estimator, the FIM is given by

$$\begin{aligned}\mathcal{F}_N(\boldsymbol{s}) &= E\{(\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}])(\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}])^T\} \\ &= -E\{\nabla_{\boldsymbol{s}}(\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}])^T\}\end{aligned} \tag{8}$$

where $p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}]$ is the conditional probability density function (pdf) of the measurements, given the shift $\boldsymbol{s}$ (the gradient $\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}]$ is sometimes termed score). Since v_{N-1} and v_N are Gaussian distributed and statistically independent, then, using Eq. (3) yields

$$p\{z_N[\boldsymbol{x}(i, j), \boldsymbol{s}] \mid \boldsymbol{s}\} \sim \mathcal{N}(F_N[\boldsymbol{x}(i, j)+\boldsymbol{s}] - F_{N-1}[\boldsymbol{x}(i, j)], 2\sigma^2) \tag{9}$$

Assume that K independent measurements are available. Since the measurement noise is a Gaussian distributed, spatially and temporally white random field, these measurements are independent, with identical variances and different means. The joint conditional density function is obtained as a product of the marginal densities, hence,

$$\begin{aligned}\log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}] = C(K, \sigma) - \frac{1}{4\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} [z_N[\boldsymbol{x}(i, j), \boldsymbol{s}] \\ -(F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}] - F_{N-1}[\boldsymbol{x}(i, j)])]^2\end{aligned} \tag{10}$$

where $C(K, \sigma)$ is a constant that does not depend on the shift $\boldsymbol{s}$. The score and the Hessian become, respectively,

$$\begin{aligned}\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}] = \frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} [z_N[\boldsymbol{x}(i, j), \boldsymbol{s}] \\ - (F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}] \\ - F_{N-1}[\boldsymbol{x}(i, j)])] \nabla_{\boldsymbol{s}} F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}]\end{aligned} \tag{11}$$

$$\begin{aligned}\nabla_{\boldsymbol{s}}\{\nabla_{\boldsymbol{s}} \log p[\boldsymbol{Z}_N^K \mid \boldsymbol{s}]\}^T = \frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \{-(\nabla_{\boldsymbol{s}} F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}]) \\ \times (\nabla_{\boldsymbol{s}} F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}])^T + [z_N[\boldsymbol{x}(i, j), \boldsymbol{s}] \\ - (F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}] \\ - F_{N-1}[\boldsymbol{x}(i, j)])] \nabla_{\boldsymbol{s}}(\nabla_{\boldsymbol{s}} F_N[\boldsymbol{x}(i, j) + \boldsymbol{s}])^T\}\end{aligned} \tag{12}$$

Taking the expected value of Eq. (12) yields the following expression for the FIM:

$$\mathcal{F}_N(\bar{s}) = \frac{1}{2\sigma^2}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)}) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)})^T \quad (13)$$

Equation (13) highlights the role of the brightness gradient as a directional measure for the information content in the picture. Thus, a dull scenery, characterized by small gradients, has a relatively low information content, and vice versa. Note that in practice, because of the discrete nature of the acquired picture, the brightness gradient may be approximated via finite differences.

Returning to the proof of the theorem, rewrite the estimator as

$$\hat{s}_{\mathrm{MSD}}(N) = \arg\min_s J_{\mathrm{MSD}}(s) \quad (14)$$

where $J_{\mathrm{MSD}}(s)$ is the cost function associated with the MSD estimator:

$$J_{\mathrm{MSD}}(s) = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(I_N[\boldsymbol{x}(i,j)+s] - I_{N-1}[\boldsymbol{x}(i,j)])^2 \quad (15)$$

Using Eq. (3) in Eq. (15) yields

$$J_{\mathrm{MSD}}(s) = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(F_N[\boldsymbol{x}(i,j)+s] - F_N[\boldsymbol{x}(i,j)+\bar{s}] + v_N[\boldsymbol{x}(i,j)+s] - v_{N-1}[\boldsymbol{x}(i,j)])^2 \quad (16)$$

Since the Hessian of the surface function $F_N[\boldsymbol{x}(i,j)+s]$ is bounded, then

$$(s-\bar{s})^T(\nabla_s\{\nabla_s F_N[\boldsymbol{x}(i,j)+s]\}^T|_{s=\bar{s}})(s-\bar{s}) \sim o(\Delta s) \quad (17)$$

where $\Delta s \triangleq s - \bar{s}$, and $o(\varepsilon)$ denotes a function satisfying $\lim_{\varepsilon\to 0}(o(\varepsilon)/\varepsilon) = 0$. A Taylor expansion of $F_N[\boldsymbol{x}(i,j)+s]$ about $F_N[\boldsymbol{x}(i,j)+\bar{s}]$ therefore yields

$$F_N[\boldsymbol{x}(i,j)+s] - F_N[\boldsymbol{x}(i,j)+\bar{s}] = (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T(s-\bar{s}) + o(\Delta s) \quad (18)$$

where $(\nabla_s F_N[\boldsymbol{x}(i,j)+s]\,|_{s=\bar{s}})$ denotes the gradient of $F_N[\boldsymbol{x}(i,j)+s]$ at $s=\bar{s}$. Assuming also that $v_N[\boldsymbol{x}(i,j)+s]$ can be approximated by $v_N[\boldsymbol{x}(i,j)+\bar{s}]$ for $s\to\bar{s}$, yields

$$J_{\mathrm{MSD}}(s) = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}[(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T(s-\bar{s}) + o(\Delta s) + v_N[\boldsymbol{x}(i,j)+\bar{s}] - v_{N-1}[\boldsymbol{x}(i,j)]]^2 \quad (19)$$

The necessary condition for the minimum of $J_{\mathrm{MSD}}(s)$ is:

$$\nabla_s J_{\mathrm{MSD}}(s) = \frac{2}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\{[(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T \times (s-\bar{s}) + o(\Delta s) + v_N[\boldsymbol{x}(i,j)+\bar{s}] - v_{N-1}[\boldsymbol{x}(i,j)]] \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}} + O(\Delta s))\} = 0 \quad (20)$$

where $O(\varepsilon)$ denotes a function satisfying $\lim_{\varepsilon\to 0}(O(\varepsilon)/\varepsilon) < \infty$. Noting that, for $s\to\bar{s}$, $o(\Delta s)$ and $O(\Delta s)$ are dominated by $(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T(s-\bar{s})$ and $\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}}$, respectively, we have

$$\left[\frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T\right] \times (s-\bar{s}) + \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(v_N[\boldsymbol{x}(i,j)+\bar{s}] - v_{N-1}[\boldsymbol{x}(i,j)]) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}}) = 0 \quad (21)$$

Define the vector

$$\boldsymbol{n}(N) := \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(v_N[\boldsymbol{x}(i,j)+\bar{s}] - v_{N-1}[\boldsymbol{x}(i,j)])(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}}) \quad (22)$$

Then, noting Eq. (13), we have

$$\hat{s}_{\mathrm{MSD}}(N) = \bar{s}(N) - \frac{n_x n_y}{2\sigma^2}\mathcal{F}_N^{-1}(\bar{s})\boldsymbol{n}(N) \quad (23)$$

Now, unbiasedness of $\hat{s}_{\mathrm{MSD}}(N)$ follows by observing that $E[\boldsymbol{n}(N)] = 0$. Moreover, defining the estimation error as

$$\tilde{s}_{\mathrm{MSD}}(N) \triangleq \bar{s}(N) - \hat{s}_{\mathrm{MSD}}(N) \quad (24)$$

and computing the covariance of the estimation error yields

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MSD}}(N)\} = \left[\frac{n_x n_y}{2\sigma^2}\right]^2 \mathcal{F}_N^{-1}(s)E[\boldsymbol{n}(N)\boldsymbol{n}^T(N)]\mathcal{F}_N^{-1}(\bar{s}) \quad (25)$$

To compute $E[\boldsymbol{n}(N)\boldsymbol{n}^T(N)]$, note that the picture noise is a spatially and temporally white random field; hence,

$$E[\boldsymbol{n}(N)\boldsymbol{n}^T(N)] = \frac{2\sigma^2}{(n_x n_y)^2}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}}) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T = \left[\frac{2\sigma^2}{n_x n_y}\right]^2 \mathcal{F}_N(\bar{s}) \quad (26)$$

Using Eq. (26) in Eq. (25) finally yields:

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MSD}}(N)\} = \mathcal{F}_N^{-1}(\bar{s}) \quad (27)$$

which completes the proof. □

As is well known, an efficient estimator can always be derived using a maximum likelihood (ML) approach.[12] The following theorem shows that the MSD estimator can also be characterized as a small error approximation of an ML estimator.

Theorem 2. Operating in the vicinity of the true shift (assuming that no loss of lock or false lock has occurred), the MSD estimator can be characterized as an ML estimator in the following sense:

$$\hat{s}_{\mathrm{ML}}(N) = \arg\max_s \log p_{\bar{s}(N)}(\boldsymbol{Z}_N^K \mid s)$$

where $p_{\bar{s}(N)}(\boldsymbol{Z}_N^K \mid s)$ denotes the joint conditional pdf of the K measurements (corresponding to the K pixels in the test window), whose functional form is obtained by assuming $s = \bar{s}(N)$ prior to acquiring the measurements.

Proof. An ordinary ML estimator satisfies

$$\hat{s}_{\mathrm{ML}}(N) = \arg\max_s \log p[\boldsymbol{Z}_N^K \mid s] \quad (28)$$

which, using Eq. (10), becomes

$$\hat{s}_{\mathrm{ML}}(N) = \arg\min_s \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}[z_N[\boldsymbol{x}(i,j),s] - (F_N[\boldsymbol{x}(i,j)+s] - F_{N-1}[\boldsymbol{x}(i,j)])]^2 \quad (29)$$

A direct minimization of the right-hand side (rhs) of Eq. (29) would require the knowledge of the unknown brightness pattern of the ground scenery, $F_{N-1}[\boldsymbol{x}(i,j)]$. However, a Taylor expansion of the conditional mean term about the true shift yields

$$F_N[\boldsymbol{x}(i,j)+s] - F_{N-1}[\boldsymbol{x}(i,j)] = (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}})^T(s-\bar{s}) + o(\Delta s) \quad (30)$$

Assume further that the search in Eq. (29) is performed over shifts satisfying $s\to\bar{s}(N)$. Since the gradient of F_N is bounded, the rhs

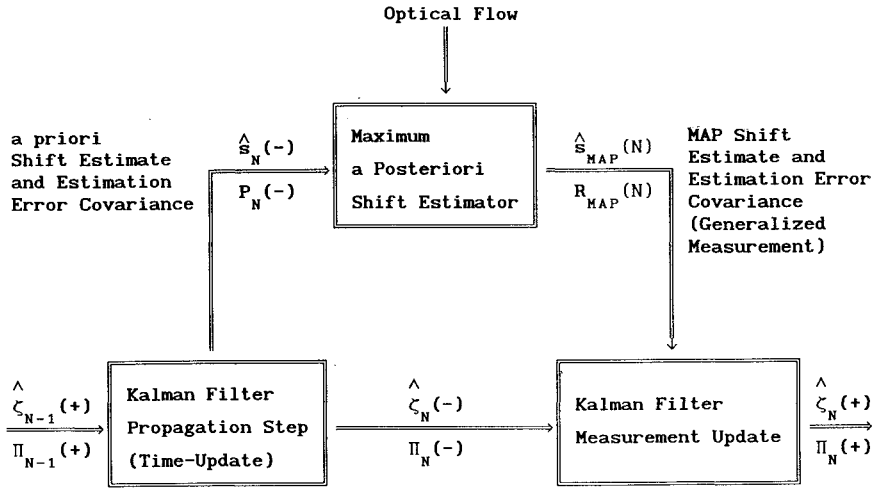


Fig. 3 MAP motion estimation $\hat{\zeta}(+)$ the a posteriori estimate, $\hat{\zeta}(-)$ the a priori estimate, $\Pi(+)$ the a posteriori estimation error covariance and $\Pi(-)$ the a priori estimation error covariance.

of Eq. (30) is $O(\Delta s)$ and, therefore, is dominated by $z_N[\boldsymbol{x}(i, j), s]$. Using Eq. (6) yields

$$\hat{s}_{\mathrm{ML}}(N) = \arg\min_{s} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (z_N[\boldsymbol{x}(i, j), s])^2 \tag{31}$$

which may be obtained directly by formally setting $s = \bar{s}(N)$ in $p[Z_N^K \mid s]$ in Eq. (28). Equation (31) results in

$$\begin{aligned}\hat{s}_{\mathrm{ML}}(N) &= \arg\min_{s} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (I_N[\boldsymbol{x}(i, j) + s] - I_{N-1}[\boldsymbol{x}(i, j)])^2 \\ &= \arg\min_{s} J_{\mathrm{MSD}}(s)\end{aligned}$$

whence $\hat{s}_{\mathrm{ML}}(N) = \hat{s}_{\mathrm{MSD}}(N)$, and the proof is completed. □

Correlation Between Consecutive Measurements

To enable the utilization of the shift estimator as a sensor in an interlaced Kalman filter/MSD estimator motion estimation scheme, its statistical characteristics have to be evaluated. The purpose of the ensuing development is to show that the shift estimation error (constituting the sensor's "measurement noise" in the overall estimation scheme) is temporally correlated and to compute that correlation.

To compute the correlation between two successive measurements, define these measurements as

$$Z_N^K \triangleq I_N[\boldsymbol{x} + \bar{s}(N)] - I_{N-1}(\boldsymbol{x}) \tag{32}$$

$$Z_{N+1}^K \triangleq I_{N+1}[\boldsymbol{x} + \bar{s}(N) + \bar{s}(N+1)] - I_N[\boldsymbol{x} + \bar{s}(N)] \tag{33}$$

i.e., it is assumed that the second frame is shifted by $\bar{s}(N)$ with respect to the first frame, and the third frame is shifted by $\bar{s}(N+1)$ with respect to the second. Notice that $I_N[\boldsymbol{x} + \bar{s}(N)]$ is used by both measurements, which brings about the correlation between them. Using Eq. (23) and noting that the MSD estimator is unbiased yields

$$\begin{aligned}&\mathrm{cov}\{\tilde{s}_{\mathrm{MSD}}(N), \tilde{s}_{\mathrm{MSD}}(N+1)\} \\ &= \left[\frac{n_x n_y}{2\sigma^2}\right]^2 \mathcal{F}_N^{-1}(\bar{s}) E[\boldsymbol{n}(N)\boldsymbol{n}^T(N+1)] \mathcal{F}_{N+1}^{-1}(\bar{s})\end{aligned} \tag{34}$$

To compute $E[\boldsymbol{n}(N)\boldsymbol{n}^T(N+1)]$, note that the statistical characteristics of the picture noise render

$$\begin{aligned}&E\{(v_N[\boldsymbol{x}(i, j) + \bar{s}(N)] - v_{N+1}[\boldsymbol{x}(i, j)]) \\ &\quad \times (v_{N+1}[\boldsymbol{x}(l, m) + \bar{s}(N+1)] - v_N[\boldsymbol{x}(l, m)])\} \\ &= \begin{cases} -\sigma^2 & \text{for } l, m \text{ such that } \boldsymbol{x}(i, j) + \bar{s} = \boldsymbol{x}(l, m) \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{35}$$

Hence, using definition (22) yields

$$\begin{aligned}&E[\boldsymbol{n}(N)\boldsymbol{n}^T(N+1)] \\ &= -\left[\frac{\sigma}{n_x n_y}\right]^2 \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (\nabla_s F_N[\boldsymbol{x}(i, j) + s]|_{s=\bar{s}(N)}) \\ &\quad \times (\nabla_s F_{N+1}[\boldsymbol{x}(i, j) + \bar{s}(N) + s]|_{s=\bar{s}(N+1)})^T\end{aligned} \tag{36}$$

Now, according to Eq. (1)

$$F_{N+1}[\boldsymbol{x}(i, j) + \bar{s}(N) + \bar{s}(N+1)] = F_N[\boldsymbol{x}(i, j) + \bar{s}(N)] \tag{37}$$

Substituting Eq. (37) into Eq. (36) yields

$$E[\boldsymbol{n}(N)\boldsymbol{n}^T(N+1)] = -2\left[\frac{\sigma^2}{n_x n_y}\right]^2 \mathcal{F}_N(\bar{s}) \tag{38}$$

Using Eq. (38) in Eq. (34) finally gives

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MSD}}(N), \tilde{s}_{\mathrm{MSD}}(N+1)\} = -\tfrac{1}{2}\mathcal{F}_{N+1}^{-1}(\bar{s}) \tag{39}$$

This correlation between measurements has to be accounted for when designing an appropriate Kalman filter.

IV. Maximum a Posteriori Image Registration

Similarly to the estimation scheme described in the previous section, the MAP image registration algorithm developed in this section is regarded as an optical flow sensor by the Kalman motion estimator. However, whereas in the previous scheme only a weak feedback existed between the Kalman filter and the MSD estimator, in the MAP algorithm the prediction statistics provided by the Kalman filter are used to smooth the estimates generated by the optical flow estimator, thus rendering an improved estimation performance and enhanced robustness.

The Fisher approach underlying the MSD estimator raised a problem, since no rigorous mechanism was provided for the optical flow estimator to utilize any prior information regarding the interframe shift, such as the prediction statistics generated by the Kalman filter. Note, in that regard, that in Ref. 2, the predicted shift computed by the time-update stage of the Kalman filter was used by the shift estimator in a heuristic manner, as an initial condition for the global search. No statistical significance was attached to this estimate.

Using a Bayesian estimation methodology, on the other hand, the interframe shift is considered a random variable, having an a priori probability density that reflects the available information regarding that variable prior to conducting the measurement. This information is provided to the shift estimator by the prediction stage of the Kalman filter. The utilization of the a priori estimate and its prediction error covariance is performed in the combined estimation scheme according to Fig. 3, where ζ denotes the state of the

total dynamic system and Π denotes the associated estimation error covariance.

Derivation

Let $\mathcal{Z}^K(N)$ denote the measurement history, that is,

$$\mathcal{Z}^K(N) \triangleq [\mathbf{Z}_1^K, \mathbf{Z}_2^K, \ldots, \mathbf{Z}_N^K] \tag{40}$$

where $\mathbf{Z}_N^K$ is the measurement vector at time N (comprising the K pixels measured in the test window). A discrete-time Kalman filter is assumed to provide the shift estimator with a priori information (predicted shift and prediction error covariance). Denote the predicted shift, after processing $N-1$ measurements, $\hat{s}_N(-)$, and its prediction error covariance $P_N(-)$. It is also assumed that the pdf of the predicted shift estimate, comprising the information rendered to the shift estimator, is Gaussian, i.e.,

$$p[s(N) \mid \mathcal{Z}^K(N-1)] \sim \mathcal{N}(\hat{s}_N(-), P_N(-)) \tag{41}$$

The MAP estimate of the shift is found via

$$\hat{s}_{\mathrm{MAP}}(N) = \arg\max_s p[s \mid \mathcal{Z}^K(N)] \tag{42}$$

Using Bayes rule in Eq. (42) yields

$$p[s \mid \mathcal{Z}^K(N)] = \frac{p[\mathbf{Z}_N^K \mid s, \mathcal{Z}^K(N-1)]p[s \mid \mathcal{Z}^K(N-1)]}{p[\mathbf{Z}_N^K \mid \mathcal{Z}^K(N-1)]} \tag{43}$$

Given the shift s, $\mathbf{Z}_N^K$ is independent of $\mathcal{Z}^K(N-1)$, hence,

$$\hat{s}_{\mathrm{MAP}}(N) = \arg\max_s (\log p[\mathbf{Z}_N^K \mid s] + \log p[s \mid \mathcal{Z}^K(N-1)] \tag{44}$$

Employing Eqs. (10) and (41) yields

$$\hat{s}_{\mathrm{MAP}}(N) = \arg\max_s \left\{ \frac{-1}{4\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} [z_N[\boldsymbol{x}(i,j), s] - (F_N[\boldsymbol{x}(i,j)+s] - F_{N-1}[\boldsymbol{x}(i,j)])]^2 - \frac{1}{2}(s-\hat{s}_N(-))^T P_N^{-1}(-)(s-\hat{s}_N(-)) \right\} \tag{45}$$

Assuming that $s \to \bar{s}(N)$ and noting that $F_N[\boldsymbol{x}(i,j)+s] - F_{N-1}[\boldsymbol{x}(i,j)]$ is dominated by $z_N[\boldsymbol{x}(i,j), s]$ (as observed in the previous section), yields

$$\hat{s}_{\mathrm{MAP}}(N) = \arg\min_s \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (I_N[\boldsymbol{x}(i,j)+s] - I_{N-1}[\boldsymbol{x}(i,j)])^2 + (s-\hat{s}_N(-))^T P_N^{-1}(-)(s-\hat{s}_N(-)) \right\} \tag{46}$$

which renders the sought for MAP algorithm.

Discussion

The actual implementation of the MAP algorithm is similar to the implementation of the MSD algorithm, except that the modified cost function (46) is used. The MAP algorithm differs from the MSD algorithm by the addition of the a priori information, which is appropriately weighted by its statistical significance. Moreover, by virtue of the numerical properties of the quadratic form, the robustness of the MAP algorithm with respect to loss of lock is considerably improved over that of the conventional MSD algorithm.

Statistical Analysis

Theorem 3. Under the assumptions underlying Theorem 1, the MAP estimator is unbiased and efficient. Moreover, its estimation error covariance is smaller than that of the MSD shift estimator. □

Proof. To assess the efficiency of the MAP shift estimation algorithm, the CRLB associated with the estimation of a random shift is computed. The information matrix for a random shift vector is given by

$$\mathcal{I}_N(s) = -E\{\nabla_s(\nabla_s \log p[\mathbf{Z}_N^K \mid s])^T\} - E\{\nabla_s(\nabla_s \log p[s \mid \mathcal{Z}^K(N-1)])^T\} \tag{47}$$

The first term on the rhs of Eq. (47) is recognized to be the FIM for an unknown but deterministic shift, Eq. (13). Evaluating the second term on the rhs for a rectangular $n_x \times n_y$ test window results in

$$\mathcal{I}_N(\bar{s}) = \frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)}) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)})^T + P_N^{-1}(-) = \mathcal{F}_N(\bar{s}) + P_N^{-1}(-) \tag{48}$$

Comparing the last result with the FIM obtained for the nonrandom shift, it is noted that, as can be expected, the addition of a priori information decreases the lower bound. Obviously, any efficient estimator will, in this case, outperform the MSD estimator.

Returning to the MAP estimator, the MAP cost function is rewritten as

$$J_{\mathrm{MAP}}(s) = \frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (I_N[\boldsymbol{x}(i,j)+s] - I_{N-1}[\boldsymbol{x}(i,j)])^2 + (s-\hat{s}_N(-))^T P_N^{-1}(-)(s-\hat{s}_N(-)) \tag{49}$$

Using Eq. (15) this cost function is related to the MSD cost function according to

$$J_{\mathrm{MAP}}(s) = \frac{n_x n_y}{2\sigma^2} J_{\mathrm{MSD}}(s) + (s-\hat{s}_N(-))^T \times P_N^{-1}(-)(s-\hat{s}_N(-)) \tag{50}$$

A Taylor expansion of the quadratic form on the rhs of Eq. (50) about the true shift $\bar{s}(N)$ yields

$$J_{\mathrm{MAP}}(s) = \frac{n_x n_y}{2\sigma^2} J_{\mathrm{MSD}}(s) + (\bar{s}(N)-\hat{s}_N(-))^T P_N^{-1}(-)(\bar{s}(N)-\hat{s}_N(-)) + 2(\bar{s}(N)-\hat{s}_N(-))^T P_N^{-1}(-)(s-\bar{s}(N)) + (s-\bar{s}(N))^T P_N^{-1}(-)(s-\bar{s}(N)) \tag{51}$$

whence

$$\nabla_s J_{\mathrm{MAP}}(s) = \frac{n_x n_y}{2\sigma^2} \nabla_s J_{\mathrm{MSD}}(s) + 2P_N^{-1}(-)(\bar{s}(N)-\hat{s}_N(-)) + 2P_N^{-1}(-)(s-\bar{s}(N)) \tag{52}$$

The necessary condition for a minimum of J_{MAP} therefore yields

$$\frac{n_x n_y}{2\sigma^2} \nabla_s J_{\mathrm{MSD}}(s) + 2P_N^{-1}(-)(\bar{s}(N)-\hat{s}_N(-)) + 2P_N^{-1}(-)(s-\bar{s}(N)) = 0 \tag{53}$$

Employing Eq. (21) yields

$$\left[\frac{1}{\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)}) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)})^T \right] (s-\bar{s}(N)) + \frac{1}{\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (v_N[\boldsymbol{x}(i,j)+\bar{s}] - v_{N-1}[\boldsymbol{x}(i,j)]) \times (\nabla_s F_N[\boldsymbol{x}(i,j)+s]|_{s=\bar{s}(N)}) + 2P_N^{-1}(-)(\bar{s}(N) - \hat{s}_N(-)) + 2P_N^{-1}(-)(s-\bar{s}(N)) = 0 \tag{54}$$

Using definitions (13) and (22) in the last equation yields

$$\hat{s}_{\mathrm{MAP}}(N) = \bar{s}(N) - (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1} \times \left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right] \tag{55}$$

where the Kalman filter prediction error is defined as

$$\tilde{s}_N(-) \triangleq \bar{s}(N) - \hat{s}_N(-)$$

Defining analogously the MAP estimation error as

$$\tilde{s}_{\mathrm{MAP}}(N) \triangleq \bar{s}(N) - \hat{s}_{\mathrm{MAP}}(N)$$

yields

$$\tilde{s}_{\mathrm{MAP}}(N) = \left(\mathcal{F}_N(\bar{s}) + P_N^{-1}(-)\right)^{-1}\left(\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right) \tag{56}$$

It follows that the MAP estimator is unbiased, since, as noted previously, $E[\boldsymbol{n}(N)] = 0$, and the Kalman predicted estimate is unbiased. Moreover, computing the covariance of the estimation error yields

$$E\{[\bar{s}(N) - \hat{s}_{\mathrm{MAP}}(N)][\bar{s}(N) - \hat{s}_{\mathrm{MAP}}(N)]^T\} = (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1}\mathrm{cov}\left\{\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right\} \times (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1} \tag{57}$$

But $\boldsymbol{n}(N)$ is independent of the prediction error $\tilde{s}_N(-)$, hence, using Eq. (26) finally yields

$$E\left\{[\bar{s}(N) - \hat{s}_{\mathrm{MAP}}(N)][\bar{s}(N) - \hat{s}_{\mathrm{MAP}}(N)]^T\right\} = (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1} \tag{58}$$

The last result shows that the MAP estimator is efficient under the theorem's assumptions, thus completing the proof. □

Correlation Between Successive Measurements

To compute the correlation between successive measurements, the procedure used in Sec. III is followed. Since the MAP estimator was shown to be unbiased, then

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MAP}}(N), \tilde{s}_{\mathrm{MAP}}(N+1)\} = E\{\tilde{s}_{\mathrm{MAP}}(N)\tilde{s}_{\mathrm{MAP}}^T(N+1)\} = (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1}E\left\{\left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right] \times \left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N+1) + P_{N+1}^{-1}(-)\tilde{s}_{N+1}(-)\right]^T\right\} \times (\mathcal{F}_{N+1}(\bar{s}) + P_{N+1}^{-1}(-))^{-1} \tag{59}$$

Employing the orthogonality of the Kalman filter prediction error $\tilde{s}_{N+1}(-)$ to $\mathcal{Z}_N^K$ yields

$$E\left\{\left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right] \times \left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N+1) + P_{N+1}^{-1}(-)\tilde{s}_{N+1}(-)\right]^T\right\} = \left[\frac{n_x n_y}{2\sigma^2}\right]^2 E[\boldsymbol{n}(N)\boldsymbol{n}^T(N+1)] + P_N^{-1}(-)E[\tilde{s}_N(-)\tilde{s}_{N+1}^T(-)]P_{N+1}^{-1}(-) \tag{60}$$

The prediction error is a first-order Markov process, with

$$E[\tilde{s}_N(-)\tilde{s}_{N+1}^T(-)] = P_N(-)[I - K_N H_N]^T\Phi_N^T \tag{61}$$

where K_N, H_N and Φ_N are, respectively, the Kalman gain, the measurement matrix and the transition matrix. Using Eqs. (38) and (61) in Eq. (60) results in

$$E\left\{\left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N) + P_N^{-1}(-)\tilde{s}_N(-)\right] \times \left[\frac{n_x n_y}{2\sigma^2}\boldsymbol{n}(N+1) + P_{N+1}^{-1}(-)\tilde{s}_{N+1}(-)\right]^T\right\} = -\frac{1}{2}\mathcal{F}_N(\bar{s}) + [I - K_N H_N]^T\Phi_N^T P_{N+1}^{-1}(-) \tag{62}$$

Substituting Eq. (62) into Eq. (59) gives

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MAP}}(N), \tilde{s}_{\mathrm{MAP}}(N+1)\} = (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1}(-\tfrac{1}{2}\mathcal{F}_N(\bar{s}) + [I - K_N H_N]^T\Phi_N^T P_{N+1}^{-1}(-))(\mathcal{F}_{N+1}(\bar{s}) + P_{N+1}^{-1}(-))^{-1} \tag{63}$$

This expression can be somewhat simplified if the following definition of C_N, the Rauch-Tung-Striebel fixed-interval smoother gain[13] is used:

$$C_N \triangleq P_N(+)\Phi_N^T P_{N+1}^{-1}(-) \tag{64}$$

Using definition (64) in Eq. (63) finally yields

$$\mathrm{cov}\{\tilde{s}_{\mathrm{MAP}}(N), \tilde{s}_{\mathrm{MAP}}(N+1)\} = (\mathcal{F}_N(\bar{s}) + P_N^{-1}(-))^{-1}(-\tfrac{1}{2}\mathcal{F}_N(\bar{s}) + P_N^{-1}(-)C_N) \times (\mathcal{F}_{N+1}(\bar{s}) + P_{N+1}^{-1}(-))^{-1} \tag{65}$$

which is the required result.

V. Experimental Investigation

An extensive experimental investigation was carried out at the Flight Control Laboratory of the Technion. The experimental program was designed 1) to investigate the properties of the estimators previously presented under conditions which are close to reality (involving unmodeled disturbances and parameter uncertainties) and 2) to demonstrate the superiority of the MAP estimation scheme over the conventional Kalman filter/MSD estimator scheme.

Experimental Setup

The experimental investigation utilized a two-axis light table, controlled by a 16 MHz Motorola 1131 real-time computer (using an MC68020/MC68881 CPU/FPU combination). The electro-optical sensor was a Javelin Charge Coupled Device camera, with an image size of 573 × 573 pixels. The image was captured by an FG100 image digitizing card, which sampled an image size of 512 × 512 pixels at a rate of 25 frames/s. The images used were real aerial photographs, representing a variety of image textures. The experimental layout is shown in Fig. 4.

The tests involved moving the table along straight lines at certain directions, thus simulating a straight and level flight of the airborne camera above the ground scenery. During the operation of the table, the image was sampled, and the digital gray level array was transmitted to a real-time computer, at a rate of 1 frame/40 ms. The motion estimation algorithms implemented in the experimental investigation were 1) the Kalman filter/MSD shift estimator scheme and 2) the Kalman filter/MAP estimation scheme. The Kalman filter implemented in both cases was designed to handle the correlated measurement noise, as will be briefly described in the following.

Kalman Filter Design

To keep the investigation tractable, a simple, constant velocity kinematic model was assumed for the camera-carrying platform. The resulting discrete-time interframe shift dynamic equation was

$$\bar{s}(k+1) = \bar{s}(k) + \boldsymbol{w}(k) \tag{66}$$

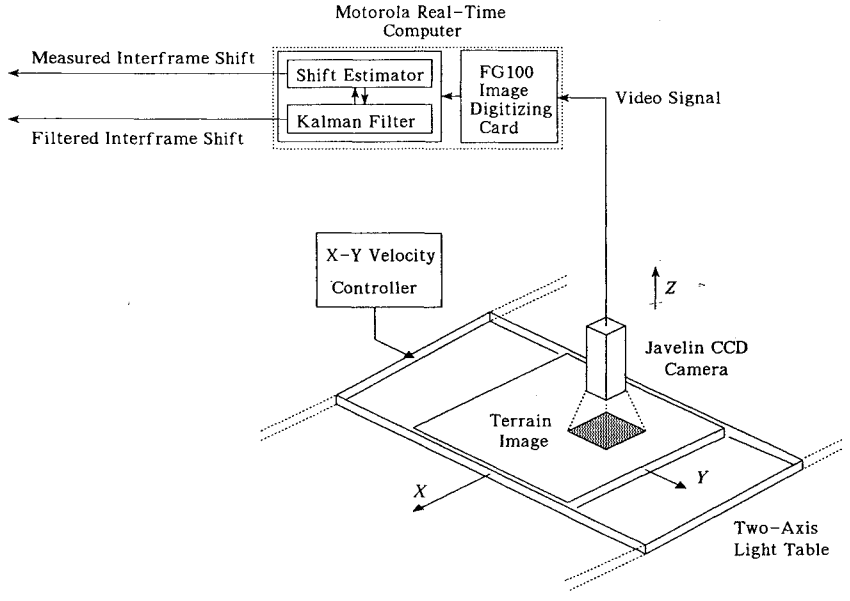


Fig. 4 Experimental layout.

where $w(k)$ is the process noise sequence (which would normally result from turbulence effects, unmodeled dynamics, etc.). The measurement equation pertaining to the shift estimator was

$$z(k+1) = \bar{s}(k+1) + e(k+1) \tag{67}$$

where $e(k)$ is the estimation error of the image registration algorithm, considered in this context as the measurement noise sequence. Since, as previously shown, the shift estimation error is a correlated sequence, it was assumed to obey the following autoregressive model equation:

$$e(k+1) = \alpha e(k) + r(k) \tag{68}$$

where $r(k)$ is a white, zero-mean Gaussian noise. Assuming stationarity, the covariance of the driving white noise r is computed as

$$\text{cov}\{r\} = (1-\alpha^2) \cdot \text{cov}\{e\} \tag{69}$$

where cov$\{e\}$, the estimation error covariance, was computed previously. The correlation coefficient α is related to the estimation error correlation (also previously computed) according to

$$\text{cov}\{e(k+1), e(k)\} = \alpha \cdot \text{cov}\{e(k)\} \tag{70}$$

A straightforward implementation of standard Kalman filtering theory was clearly not possible, because of the inherently correlated measurement noise. Therefore, the following procedure was adopted to design an appropriate filtering algorithm.[14]

1) The system state was augmented, to include the shift estimation error e.

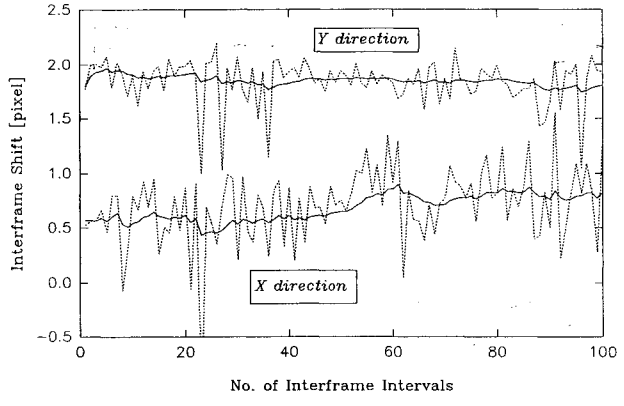
2) Since the resulting augmented model had a singular measurement, a reduced-order observer was designed by transforming the augmented state. The resulting transformed measurement noise was correlated with the transformed process noise.

3) A special filtering algorithm, whose time-update level takes into account the correlation between the process noise and the measurement noise was then implemented for the reduced-order model.

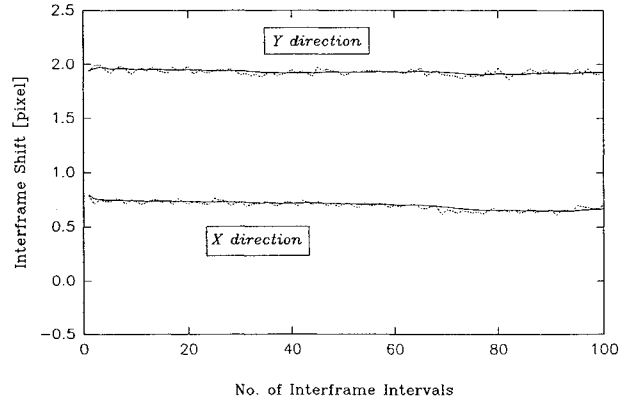
4) The parameters of the resulting filter were tuned using a hypothesis testing technique,[15] yielding a statistically consistent filter.

Results

Numerous experiments were performed to investigate the statistical properties and the performance of the shift estimation algorithms under various conditions. The dependence of the estimation performance on several parameters was examined. The investigation variables included the test window size and shape, and the image resolution and texture (which directly affected the brightness gradients). Typical results are shown in Figs. 5 and 6, which are presented to qualitatively demonstrate the superior performance of the MAP algorithm. In these figures, the shift measurement (i.e., the output of the image registration algorithm) and the filtered shift estimate (i.e.,

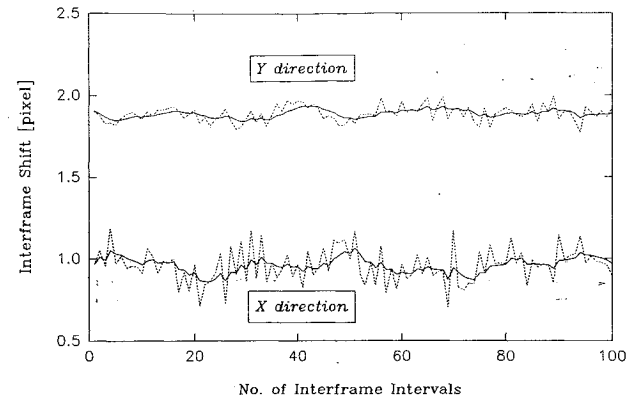


a)

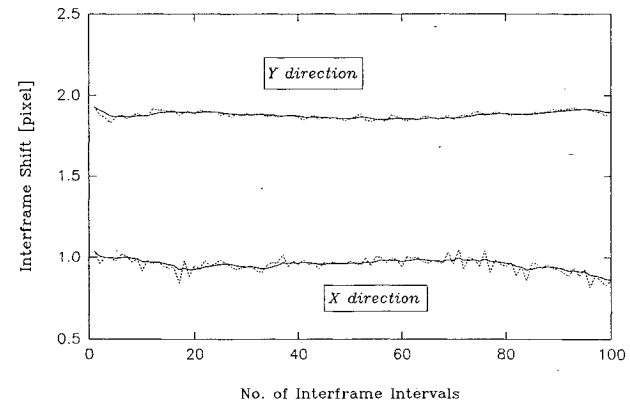


b)

Fig. 5 Experimental results for dull-texture scenery with measured (dotted line) and filtered (solid line) shift: a) MSD algorithm and b) MAP algorithm.



a)



b)

Fig. 6 Experimental results for rich-texture scenery with measured (dotted line) and filtered (solid line) shift: a) MSD algorithm and b) MAP algorithm.

the output of the Kalman filter) are presented for two different image textures (i.e., two different ground sceneries). In both experiments, the motion of the light table was programmed to simulate a constant velocity/constant altitude straight and level flight. The nominal programmed motion of the light table was contaminated by a stochastic process noise (possessing unknown statistics), generated by the nonideal actuator of the light table. The results corresponding to a desert area, having a dull texture, are shown in Fig. 5. As can be observed, the MSD shift measurement was rather noisy, having an error standard deviation of about 0.4 pixel and experiencing several occasions of loss of lock. Nevertheless, the relatively high measurement noise generated by the MSD algorithm was filtered out quite effectively by the Kalman filter, lowering the error standard deviation to about 0.1 pixel.

On the other hand, the MAP shift estimation algorithm performed much better than the MSD algorithm. Its measurement error standard deviation was about 0.04 pixel, and no loss of lock was observed. The estimation error standard deviation was about 0.03 pixel. As can be expected, the performance of the MSD algorithm improved considerably in a rich-texture scenery, as shown in Fig. 6, yielding measurement noise and estimation error standard deviations ofabout 0.12 and 0.07 pixels, respectively. However, the MAP algorithm was still clearly superior to the MSD algorithm, resulting in measurement noise and estimation error standard deviations of 0.035 and 0.025 pixels, respectively.

VI. Conclusions

A novel method has been presented for on-line, recursive image registration/motion estimation, based on a sequence of terrain images acquired by an airborne down-looking electro-optical sensor. The method consists of an interlaced computational scheme, comprising a Kalman motion estimator and a new MAP image registration algorithm. The MAP algorithm draws statistical information from the prediction level of the Kalman filter, thus forming a robustified image registration algorithm. The conventional MSD algorithm was shown to be unbiased and statistically efficient, under reasonable conditions. Moreover, it was shown that the ordinary MSD algorithm can be characterized as a small error approximation of a maximum likelihood estimator. The new MAP algorithm was proved to be unbiased and efficient, and its estimation error covariance was shown to be smaller than that of the ordinary MSD algorithm. An experimental investigation using real aerial photographs demonstrated the superior robustness of the new MAP algorithm over that of the MSD method.

In closing we note that, although motivated by an aerospace application, the new method should also be attractive in robotics and computer vision applications.

Acknowledgments

This research was supported by the Department of Research and Development of the Israeli Ministry of Defense and by TAMAM—Precision Instruments Industries, a subsidiary of Israel Aircraft Industries. Helpful discussions with Professor S. Merhav of the Faculty of Aerospace Engineering are gratefully acknowledged.

References

[1]Bresler, Y., and Merhav, S. J., "On-Line Vehicle Motion Estimation from Visual Terrain Information, Part II: Ground Velocity and Position Estimation,"*IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-22, No. 5, 1986, pp. 588–604.

[2]Topaz, L., and Grunwald, A. J., "Enhanced Orbital Gyrocompassing by the Optical Flow Sensed by an Earth Pointing Camera," *Proceedings of the 32nd Israel Annual Conference on Aviation and Astronautics*, (Israel), Feb. 1992, pp. 229–242.

[3]Heel, J., "Dynamic Motion Vision," *Robotics and Autonomous Systems*, Vol. 6, 1990, pp. 297–314.

[4]Bresler, Y., and Merhav, S. J., "Recursive Image Registration with Application to Motion Estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-35, No. 1, 1987, pp. 70–85.

[5]Kearney, J. K., Thompson, W. B., and Boley, D. L., "Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 2, 1987, pp. 229–244.

[6]Aggarwal, J. K., and Nandhakumar, N., "On the Computation of Motion from Sequences of Images—A Review," *Proceedings of the IEEE*, Vol. 76, No. 8, 1988, pp. 917–935.

[7]Matthies, L., Kanade, T., and Szeliski, R., "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences," *International Journal of Computer Vision*, Vol. 3, 1989, pp. 209–236.

[8]Sridhar, B., Cheng, V. H. L., and Phatak, A. V., "Kalman Filter Based Range Estimation for Autonomous Navigation Using Imaging Sensors," *Proceedings of XIth International Federation of Automatic Control Symposium on Automatic Control in Aerospace*, Tsukuba, Japan, 1989.

[9]Mostafavi, H., and Smith, F. W., "Image Correlation with Geometric Distortion; Part I: Acquisition Performance," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-14, No. 3, 1978, pp. 487–493.

[10]Bar-Shalom, Y., Shertukde, H. M., and Pattipati, K. R., "Use of Measurements from an Imaging Sensor for Precision Target Tracking," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-25, No. 6, 1989, pp. 863–871.

[11]Schweppe, F. C., *Uncertain Dynamic Systems*, Prentice-Hall, Englewood-Cliffs, NJ, 1973.

[12]Sorenson, H. W., *Parameter Estimation: Principles and Problems*, Marcel Dekker, New York, 1980, pp. 185, 186.

[13]Rauch, H. E., Tung, F., and Striebel, C. T., "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal*, Vol. 3, No. 8, 1965, pp. 1445–1450.

[14]Mendel, J. M., *Lessons in Digital Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1987, pp. 225–233.

[15]Bar-Shalom, Y. and Fortmann, T. E., *Tracking and Data Association*, Academic Press, New York, 1988.