Simultaneous Motion Estimation and Filtering of Image Sequences

Chieh-Min Fan, Member, IEEE, and Nader M. Namazi, Senior Member, IEEE

Abstract—In this paper, we present an algorithm to simultaneously estimate multi-frame motions and filter image sequences. The relative motion \( \delta_k(x) \) between the reference frame \( s(x) \) and the \( k \)-th frame \( s(x - \delta_k(x)) \) in the presence of additive white Gaussian noise (AWGN) is used to estimate the maximum-likelihood (ML) principle. The reference frame is also filtered in the linear minimum mean square error (LMMSE) sense during the process of motion estimation. Simulation experiments are performed using the affine motion model to illustrate the performance of this method.

Index Terms—Image sequence filtering, maximum likelihood estimation, motion estimation.

I. INTRODUCTION

THE USE of digital video has dramatically increased recently in diverse fields such as entertainment, visual communication, multimedia, education, medicine, surveillance, remote control, and scientific research [1]. Applications exist in television, target tracking, robot navigation, dynamic monitoring of industrial processes, study of cell motion by microscopy, and highway traffic control. However, during the acquisition, transmission and storage of the image sequences, they are often corrupted by noise. Noise in a sequence not only degrades the visual quality, but also hinders the subsequent analysis and processing (e.g., compression and coding). A technique to restore the corrupted sequences is needed. The problem of removing noise from image sequences has attracted a number of researchers [2]–[7]. In this paper, we have extended Stuller’s [8] and Shatla’s [9] works to simultaneously estimate image motion field and filter noisy image sequences. Specifically, we have generalized Stuller’s method from one-dimensional (1-D), single time delay problem to two-dimensional (2-D), multiple frame motion estimation. The advantages of this method are two-fold: simultaneously estimating image motion field and filtering noisy image sequences. The intent is to simultaneously estimate \( s(x) \) and \( \delta_k(x) \) for \( k = 2, 3, \cdots, P \). In the following derivations we assume that the first frame is the reference frame.

The motion \( \delta_k(x) \) can be expanded by any convenient set of scalar, complete and orthonormal (CON) basis functions \( \{ \psi_{ij}(x) \} \), i.e.,

\[
\delta_k(x) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} d_{kij} \psi_{ij}(x)
\]

where \( x = [x_1, x_2]^T \) is the spatial vector, \( T \) denotes vector/matrix transposition, \( r_k(x) \) is the \( k \)-th observed image, \( k \) is the frame index, and \( s(x) \) is the intensity of the image which is assumed to be zero-mean Gaussian with known covariance function \( K_s(x, u) \). In addition, \( s(x - \delta_k(x)) \) are the noise-free displaced images, \( d_k(x) = [d_{k1}(x) d_{k2}(x)]^T \) are the displacement vector functions with respect to the reference frame, and \( u_k(x) \)’s are assumed to be stationary zero-mean white Gaussian noises with four sided spectral density \( N_0/4 \) and covariance function \( \xi[u_k(x)u_k(u)] = (N_0/4) \delta(x - u) \), where \( \xi[\cdot] \) denotes the expectation and \( \delta(\cdot) \) is the impulse function. The noise fields and the image process \( s(x) \) are assumed to be mutually uncorrelated. This model does not take into account changes in scene lighting conditions or translation/rotation in a plane perpendicular to that of the image frame. It is also assumed that no occlusion occurs. If the first frame is chosen as the reference frame, then

\[
\delta_1(x) = 0.
\]

The intent is to simultaneously estimate \( s(x) \) and \( \delta_k(x) \) for \( k = 2, 3, \cdots, P \). In the following derivations we assume that the first frame is the reference frame.

This paper is organized as follows. In Section II, we present the derivation of the log-likelihood function. In Section III, we construct a linear minimum mean square error (LMMSE) filter to filter the reference frame. Section IV deals with the implementation of the algorithm. In Section V, we present experimental results. Finally, the summary is given in Section VI.

II. DERIVATION OF THE LOG-LIKELIHOOD FUNCTION

In this section, we closely follow Stuller’s derivation [8] to derive the log-likelihood function. The motion \( \delta_k(x) \) can be expanded by any convenient set of scalar, complete and orthonormal (CON) basis functions \( \{ \psi_{ij}(x) \} \), i.e.,

\[
d_k(x) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} d_{kij} \psi_{ij}(x)
\]

where \( d_{kij} = [d_{kij}^{(1)} d_{kij}^{(2)}]^T \) can be obtained from

\[
d_{kij} = \int_{\Omega} \psi_{ij}(x) d_k(x) dx
\]

where \( \Omega \) is the support of \( d_k(x) \). Arranging the coefficients \( d_{kij} \) in a column-ordered vector denoted \( \delta_k \), then the noise-free
images can be written as
\[ s(x - d_k(x)) = s(x; \delta_k), \quad k = 1, 2, \ldots, P. \] (5)

Hence, the image sequence in (1) can be expressed in vector form as
\[ r(x) = \begin{bmatrix} \cdots \cr \cdots \cr s(x; \delta_2) \cr \cdots \cr s(x; \delta_P) \cr \cdots \cr s(x, 0, \delta_2, \ldots, \delta_P) \end{bmatrix} + \begin{bmatrix} \cdots \cr \cdots \cr w_1(x) \cr \cdots \cr w_2(x) \cr \cdots \cr w_P(x) \end{bmatrix} = s(x, 0, \delta_2, \ldots, \delta_P) + w(x) \] (6)

where \( r(x) = [r_1(x), r_2(x), \ldots, r_P(x)]^T \).

If \( d_k(x) \) are known to be \( D_k(x), \quad k = 2, \ldots, P, \) then the fields \( s(x) \) and \( s(x - D_k(x)), \quad k = 2, 3, \ldots, P, \) are zero-mean jointly Gaussian. By definition, the noise fields \( u_k(x), \quad k = 1, 2, \ldots, P, \) are zero-mean and Gaussian. Assuming that \( \Delta_k \) be the vector of coefficients corresponding to the motion fields \( D_k(x), \) then \( r(x) \) is also zero-mean and jointly Gaussian vector field with the following \( P \times P \) matrix covariance function
\[ K_{r \in \{ \delta_2, \ldots, \delta_P \}}(x, u; \Delta_2, \ldots, \Delta_P) = \mathbb{E}\{r(x)^T(u) | \Delta_2 = \Delta_2, \ldots, \Delta_P = \Delta_P\} = \mathbb{E}\{s(x; \Delta_2, \ldots, \Delta_P)^T(u; \Delta_2, \ldots, \Delta_P)\} + \mathbb{E}\{w(x)^T w(u)\} = K_{s \in \{ \delta_2, \ldots, \delta_P \}}(x, u; \Delta_2, \ldots, \Delta_P) + \frac{N_0}{4} \delta(x - u) \] (7)

where \( I \) is the \( P \times P \) identity matrix, and \( \delta(\cdot) \) is the impulse function.

The generalized Karhunen–Loeve (KL) expansion of \( r(x) \) can be written as
\[ r(x) = \lim_{M \to \infty} \sum_{i=1}^{M} \sum_{j=1}^{P} r_{ij} \Phi_{ij}(x; \Delta_2, \ldots, \Delta_P) \] (8)

where
\[ r_{ij} = \int_{\Gamma} \Phi_{ij}^T(x; \Delta_2, \ldots, \Delta_P) r(x) \, dx \] (9)

\( \Phi_{ij}(x; \Delta_2, \ldots, \Delta_P) \) are the \( P \times 1 \) normalized vector eigenfunctions of the covariance function matrix \( K_{s \in \{ \delta_2, \ldots, \delta_P \}}(x, u; \Delta_2, \ldots, \Delta_P), \) and \( \Gamma \) is the support of the image. We assume that \( \{ \Phi_{ij}(x; \Delta_2, \ldots, \Delta_P) \} \) is complete.

It can be shown that if the values of \( \Delta_k \) are equal to the true values of \( \delta_k, \) then the parameters \( r_{ij} \) are statistically independent Gaussian random variables having zero-mean and variances given by [8]
\[ \mathbb{E}\{r_{ij}^2 | \delta_2 = \Delta_2, \ldots, \delta_P = \Delta_P\} = \lambda_{ij}(\Delta_2, \ldots, \Delta_P) + \frac{N_0}{4} \] (10)

where \( \lambda_{ij}(\Delta_2, \ldots, \Delta_P) \) is the scalar eigenvalue corresponding to \( \Phi_{ij}(x; \Delta_2, \ldots, \Delta_P). \)

The log-likelihood function associated with \( r(x) \) is obtained by taking the logarithm of the joint probability density function of \( r_{ij} \) conditioned on \( \delta_k = \Delta_k, k = 2, 3, \ldots, P. \) It can be shown that the log-likelihood function can be expressed as [8]
\[ \ln \Lambda(\Delta_2, \ldots, \Delta_P) = I_R(\Delta_2, \ldots, \Delta_P) + I_B(\Delta_2, \ldots, \Delta_P) \] (11)

where
\[ I_R(\Delta_2, \ldots, \Delta_P) = \frac{2}{N_0} \int_{\Gamma} \int_{\Gamma} R^T(x) \cdot H_n(x, u; \Delta_2, \ldots, \Delta_P) R(u) \, du \, dx \] (12)

where \( R(x) \) is a sample vector field of \( r(x), \) and \( H_n(x, u; \Delta_2, \ldots, \Delta_P) \) is a \( P \times P \) matrix
\[ H_n(x, u; \Delta_2, \ldots, \Delta_P) = \frac{2}{N_0} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\lambda_{ij}(\Delta_2, \ldots, \Delta_P)}{\lambda_{ij}(\Delta_2, \ldots, \Delta_P) + \frac{N_0}{4}} \cdot \Phi_{ij}(x; \Delta_2, \ldots, \Delta_P) \Phi_{ij}^T(u; \Delta_2, \ldots, \Delta_P) \] (13)

that satisfies the matrix integral equation
\[ K_{s \in \{ \delta_2, \ldots, \delta_P \}}(x, u; \Delta_2, \ldots, \Delta_P) - \frac{N_0}{4} I \cdot H_n(x, u; \Delta_2, \ldots, \Delta_P) \int_{\Gamma} H_n(x, z; \Delta_2, \ldots, \Delta_P) K_{s \in \{ \delta_2, \ldots, \delta_P \}}(z, u; \Delta_2, \ldots, \Delta_P) \, dz = 0; \quad x, u \in \Gamma. \] (14)

As reported in [10] the filter \( H_n(x, u; \Delta_2, \ldots, \Delta_P) \) is the optimum filter in the LMMSE sense given \( \delta_k = \Delta_k, k = 2, 3, \ldots, P. \)

In addition, if we write (12) as
\[ I_R(\Delta_2, \ldots, \Delta_P) = \frac{2}{N_0} \int_{\Gamma} R^T(x) \hat{s}_n(x; \Delta_2, \ldots, \Delta_P) \, dx \] (15)

where
\[ \hat{s}_n(x; \Delta_2, \ldots, \Delta_P) = \int_{\Gamma} H_n(x, u; \Delta_2, \ldots, \Delta_P) \cdot R(u) \, du \] (16)

then the \( P \times 1 \) vector \( \hat{s}_n(x; \Delta_2, \ldots, \Delta_P) \) can be interpreted as the output of the noncausal filter \( H_n(x, u; \Delta_2, \ldots, \Delta_P) \) with an input vector \( R(x). \)

It can also be shown [8] that the term \( I_B(\Delta_2, \ldots, \Delta_P) \) satisfies the relation
\[ I_B(\Delta_2, \ldots, \Delta_P) = -\frac{1}{2} \int_{\Gamma} \text{Tr}[H_n(x, x; \Delta_2, \ldots, \Delta_P)] \, dx \] (17)

where \( H_n(x, x; \Delta_2, \ldots, \Delta_P) \) is the causal LMMSE estimator of \( s(x) \) from \( r(x), \) and \( \text{Tr}[\cdot] \) is the trace of a matrix.
III. THE MATRIX FILTER $H_n(x, u; \Delta_2, \cdots, \Delta_P)$

It is relatively difficult to obtain the matrix filter $H_n(x, u; \Delta_2, \cdots, \Delta_P)$ by solving (14). In this section, we generalize the method in [8] for the construction of $H_n(x, u; \Delta_2, \cdots, \Delta_P)$.

Assuming hypothetical vector functions $f_k(x) = x - D_k(x)$, $k = 2, 3, \cdots, P$, we find the inverses of $f_k(x)$ denoted $\beta_k(x)$. Then $s_k(\beta_k(x))$ is the motion compensated version of $s(x)$ from $s_k(x)$, where $s_k(x) = s(x - d_k(x))$. We perform a transformation on $r'(x)$ for $x \in \Gamma$ which results in the new $r'(u)$ for $u \in \Gamma'$. The area $\Gamma'$ consists of two parts $\Gamma_2$ and $\Gamma_2'$; the area $\Gamma_1$ is $\{u | u \in \bigcap \nabla_1(\beta_1(u)) \cap \nabla_2(\beta_2(u)) \cap \cdots \cap \nabla_P(\beta_P(u)) \}$ and the area $\Gamma_2$ is $\{u | u \in \bigcap \nabla_1(\beta_1(u)) \cap \nabla_2(\beta_2(u)) \cap \cdots \cap \nabla_P(\beta_P(u)) \}$ but $\not \in \Gamma_1$. The area $\Gamma_2$ can be further divided into many subregions. For example, $\{u | u \in \nabla_1(\beta_1(u)) \cap \nabla_2(\beta_2(u)) \cap \cdots \cap \nabla_P(\beta_P(u)) \}$, etc., where $c$ is the complement. In general, for an image sequence with $P$ frames, the area $\Gamma'$ can be divided up to $2^P - 1$ regions. To illustrate this, we overlap $r_1(u), r_2(\beta_2(u)), \cdots, r_P(\beta_P(u))$ in Fig. 1 for the case that $P = 3$. It can be seen that there are 7 regions. Then $\Gamma_1 = R_7$ and $\Gamma_2 = \{R_1 \cup R_2 \cup R_3 \cup R_4 \cup R_5 \cup R_6 \cup R_7 \}$ for this case.

In the event for which $P = 3$, we have the following transformation for each region:

$$r'(u) = \begin{bmatrix} r_1(u) \\ 0 \end{bmatrix}; \quad u \in R_2, \quad (18)$$

$$r'(u) = \begin{bmatrix} r_3(\beta_3(u)) \\ 0 \end{bmatrix}; \quad u \in R_3, \quad (19)$$

$$r'(u) = \begin{bmatrix} 2r_1(u) + r_2(\beta_2(u)) \\ 2r_1(u) - r_2(\beta_2(u)) \end{bmatrix}; \quad u \in R_4, \quad (20)$$

$$r'(u) = \begin{bmatrix} r_1(u) + r_3(\beta_3(u)) \\ r_1(u) - r_3(\beta_3(u)) \end{bmatrix}; \quad u \in R_5, \quad (21)$$

$$r'(u) = \begin{bmatrix} 2r_1(u) + r_2(\beta_2(u)) + r_3(\beta_3(u)) \\ 2r_1(u) - r_2(\beta_2(u)) - r_3(\beta_3(u)) \end{bmatrix}; \quad u \in R_6, \quad (22)$$

$$r'(u) = \begin{bmatrix} 2r_1(u) + r_2(\beta_2(u)) + r_3(\beta_3(u)) \\ 2r_1(u) - r_2(\beta_2(u)) - r_3(\beta_3(u)) \end{bmatrix}; \quad u \in R_7. \quad (23)$$

Combining (18)–(24), we have

$$r'(u) = \begin{bmatrix} s(u) \\ 0 \end{bmatrix} + \begin{bmatrix} n_2(\beta_2(u)) \\ n_2(\beta_2(u)) \end{bmatrix} \quad (25)$$

where

$$\begin{bmatrix} n_1(u) \\ n_2(u) \end{bmatrix} = \begin{bmatrix} u_1(u) \\ u_2(\beta_2(u)) \\ 0 \end{bmatrix}; \quad u \in R_2, \quad (26)$$

$$\begin{bmatrix} u_1(u) + u_2(\beta_2(u)) \\ u_1(u) - u_2(\beta_2(u)) \end{bmatrix}; \quad u \in R_4, \quad (27)$$

$$\begin{bmatrix} 2u_1(u) + 2u_2(\beta_2(u)) + u_3(\beta_3(u)) \\ 2u_1(u) - 2u_2(\beta_2(u)) - u_3(\beta_3(u)) \end{bmatrix}; \quad u \in R_7. \quad (28)$$

It can be shown that the output of the filter $H_n(x, u; \Delta_2, \Delta_3)$ is

$$\hat{s}_n(x) = [\hat{s}_1(x) \hat{s}_2(x) \hat{s}_3(x)]^T \quad (29)$$

where

$$\hat{s}_2(x) = \int_{\Gamma} g_n(x; \Delta_2, \Delta_3) [r_1'(u) - \hat{n}_1(u)] \, du \quad (30)$$

in which $g_n(x; \Delta_2, \Delta_3)$ is the impulse response function of the noncausal LMMSE point estimator $\hat{s}_n(x)$ of $s(x)$ from $s(x + n_2(\beta_2(u))) - \hat{n}_1(u)$, and $\hat{n}_1(u)$ is the LMMSE estimate of $n_1(u)$ from $n_2(u)$, i.e.,

$$\hat{n}_1(u) = \int_{\Gamma} \xi(t; u; \Delta_2, \Delta_3) n_2(t) \, dt \quad (31)$$

where $\xi(t; u; \Delta_2, \Delta_3)$ is the impulse response function of the noncausal point LMMSE estimator. Using the definition of LMMSE, the impulse response function $\xi(t; u; \Delta_2, \Delta_3)$ can be found to be

$$\xi(t; u; \Delta_2, \Delta_3) = \eta(u)[s(t - u) \quad (32)$$

Authorized licensed use limited to: NRL. Downloaded on June 11,2010 at 20:06:47 UTC from IEEE Xplore. Restrictions apply.
where
\[
\eta(u) = \begin{cases} 
0; & u \in R_1, \\
0; & u \in R_2, \\
0; & u \in R_3, \\
1 - \frac{1}{|\beta_2(u)|}; & u \in R_4, \\
1 + \frac{1}{|\beta_2(u)|}; & u \in R_5, \\
1 - \frac{1}{|\beta_2(u)|}; & u \in R_6, \\
1 + \frac{1}{|\beta_2(u)|}; & u \in R_7, \\
\end{cases}
\]
(31)

where the dot denotes the gradient of a function. In addition, \(\hat{s}_2(x)\) and \(\hat{s}_3(x)\) in (27) can be obtained from
\[
\hat{s}_k(x) = \hat{s}_1(x - D_k(x)), \quad k = 2, 3, \quad (32)
\]

It can be shown that \(\hat{s}_k(x)\) is the LMMSE estimate of \(s(x - D_k(x))\) if \(\hat{s}_k = \Lambda_k, k = 2, 3\) [8].

Using the fact that \(g_n(x, u; A_2, A_3)\) is the impulse response function of the noncausal LMMSE point estimator \(\hat{s}_n(x)\) of \(s(x)\) from \(\hat{s}_1(x) + r_1(u) - \hat{r}_1(u)\), it can be shown that \(g_n(x, u; A_2, A_3)\) is the solution to the integral equation
\[
K_n(x, u) = \int_\Gamma g_n(x, t; A_2, A_3)K_n(t, u) \, dt \\
+ g_n(x, u; A_2, A_3)Q(u)
\]
(33)

where
\[
Q(u) = \begin{cases} 
\frac{N_0}{4}; & u \in R_1, \\
\frac{N_0}{4} \frac{1}{|\beta_2(u)|}; & u \in R_2, \\
\frac{N_0}{4} \frac{1}{|\beta_2(u)|}; & u \in R_3, \\
\frac{N_0}{4} \frac{1}{1 + |\beta_2(u)|}; & u \in R_4, \\
\frac{N_0}{4} \frac{1}{1 + |\beta_2(u)|}; & u \in R_5, \\
\frac{N_0}{4} \frac{|\beta_2(u)|}{|\beta_3(u)|}; & u \in R_6, \\
\frac{N_0}{4} \frac{1}{|\beta_2(u)| + |\beta_3(u)|}; & u \in R_7, \\
\end{cases}
\]
(34)

The term \(l_R(D_2, \cdots, D_P)\) in (15) for the case \(P = 3\) is
\[
l_R(D_2, D_3) = \frac{2}{N_0} \left\{ \int_\Gamma \hat{s}_1(x)r_1(x) \, dx + \int_\Gamma \hat{s}_2(x)r_2(x) \, dx \\
+ \int_\Gamma \hat{s}_3(x)r_3(x) \, dx \right\}.
\]
(35)

Application of (18)–(24), (28), (30), and (32) into (35), results
\[
\frac{N_0}{2}l_R(D_2, D_3)
\]
\[
= \sum_{i=1}^3 \int_\Gamma \sum_{j=1}^3 \left\{ \int_{R_i} g_n(x, u; D_2, D_3)r_1(u) \, du \right\} \\
\cdot r_j(x) \, dx
\]
\[
+ \int_\Gamma \sum_{i=1}^3 \left\{ \int_{R_i} g_n(x, u; D_2, D_3)r_2(\beta_2(u)) \, du \right\} \\
\cdot r_j(x) \, dx
\]
\[
+ \int_\Gamma \sum_{i=1}^3 \left\{ \int_{R_i} g_n(x, u; D_2, D_3)r_3(\beta_3(u)) \, du \right\} \\
\cdot r_j(x) \, dx
\]
\[
+ \int_\Gamma \left\{ \int_{R_i} g_n(x, u; D_2, D_3) \right\} \\
\cdot \left[ \frac{1}{2} [1 - \eta(u)]r_1(u) \right] \, dx
\]
\[
+ \int_\Gamma \left\{ \int_{R_i} g_n(x, u; D_2, D_3) \right\} \\
\cdot \left[ \frac{1}{2} [1 + \eta(u)]r_2(\beta_2(u)) \right] \, dx
\]
\[
+ \int_\Gamma \left\{ \int_{R_i} g_n(x, u; D_2, D_3) \right\} \\
\cdot \left[ \frac{1}{2} [1 + \eta(u)]r_3(\beta_3(u)) \right] \, dx
\]
Next, we generalize the derivation above for an image sequence with \( P \) frames. It is tedious to show the transformation for each of the \( P \) regions. Expression (37), shown at the bottom of the page, illustrates the transformation for the area \( \Gamma_1 \). It can be shown that the output of the filter \( H_n(x, u; \Delta_2, \ldots, \Delta_P) \) is

\[
\hat{s}_n(x) = \left[ \hat{s}_1(x) \hat{s}_2(x) \cdots \hat{s}_P(x) \right]^T
\]

where

\[
\hat{s}_1(x) = \int_{\Gamma_1} g_n(x, u; \Delta_2, \cdots, \Delta_P) [r_1'(u) - \eta(u)] r_2'(u) du
\]

(39)

\[
g_n(x, u; \Delta_2, \cdots, \Delta_P) \quad \text{is the impulse response of the non-causal LMMSE point estimator } \hat{s}_1(x) \text{ of } s(x) \text{ from } r_1'(u) - \eta(u) r_2'(u).
\]

It can be shown that \( \delta_k(x) \) is the LMMSE estimate of \( s(x - d_k(x)) \) if \( \delta_k = \Delta_k \).

### Table I: Simulation Results

<table>
<thead>
<tr>
<th>( P )</th>
<th>SNR (dB)</th>
<th>Steady State NMSE, ( k = 2, \ldots, P )</th>
<th>ISNR (dB) for the reference frame</th>
<th>Upper Bound of ISNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>0.0003</td>
<td>7.4279</td>
<td>7.5287</td>
</tr>
<tr>
<td>10</td>
<td>0.0001</td>
<td>5.2133</td>
<td>5.3105</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.0002</td>
<td>8.1464</td>
<td>8.2822</td>
</tr>
<tr>
<td>10</td>
<td>0.0008</td>
<td>5.6716</td>
<td>5.7971</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.0030</td>
<td>8.6829</td>
<td>9.0247</td>
</tr>
<tr>
<td>10</td>
<td>0.0029</td>
<td>0.0010</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0031</td>
<td>5.7252</td>
<td>6.1250</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.0023</td>
<td>0.0010</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

\[
Q(u) = \frac{1}{4P-1} + \frac{1}{\beta_2(u)} + \cdots + \frac{1}{\beta_P(u)}; \quad u \in \Gamma_1.
\]

(41)

and

\[
\eta(u) = \frac{2^{P-1} - 1}{\beta_2(u)} + \cdots + \frac{1}{\beta_P(u)}; \quad u \in \Gamma_1.
\]

(42)

In addition

\[
\hat{s}_k(x) = \hat{s}_1(x - d_k(x)), \quad k = 2, 3, \ldots, P.
\]

(43)

It can be shown that \( \hat{s}_k(x) \) is the LMMSE estimate of \( s(x - d_k(x)) \) if \( \delta_k = \Delta_k \).

### IV. Implementation of the Algorithm

It can be seen from (15) and (39) that computing \( I_R(\Delta_2, \cdots, \Delta_P) \) is a formidable and tedious task. To compute \( I_R(\Delta_2, \cdots, \Delta_P) \), the areas of integration which are functions of \( \Delta_k \) must be found first. In addition, finding the term \( g_n(x, u; \Delta_2, \cdots, \Delta_P) \) is not trivial. Therefore, some simplifying assumptions have to be made to reduce the complexity. We assume that the motions \( d_k(x) \) are small such that few pixels move out of (or into) the images between the frame exposure. Then the contribution of the area \( \Gamma_2 \) to (39)
Fig. 3. Noisy image sequence (SNR = 10 dB, \( P = 3 \)).

Fig. 4. Filtered image sequence (SNR = 10 dB, \( P = 3 \)).

is relatively small. Therefore, by neglecting the integration over the area \( \Gamma_2 \), \( \delta_1(\mathbf{x}) \) is approximated by

\[
\delta_1(\mathbf{x}) \approx \int_{\Gamma_1} g_n(\mathbf{x}, \mathbf{u}; \Delta_2, \cdots, \Delta_P)[r'_2(\mathbf{u}) - \gamma(\mathbf{u})r'_2(\mathbf{u})] d\mathbf{u}.
\]

Consequently, \( \tilde{I}_R(\Delta_2, \cdots, \Delta_P) \) in (15) is approximated by

\[
\tilde{I}_R(\Delta_2, \cdots, \Delta_P) = \frac{2}{N_0} \int_{\Gamma_1} \mathbf{R}^T(\mathbf{x})\tilde{s}_n(\mathbf{x}; \Delta_2, \cdots, \Delta_P) d\mathbf{x}.
\]

It is further assumed that \( \Gamma_1 \) is a squared region to make the integration feasible. We also assume that \( s(\mathbf{x}) \) is a statistically stationary random field such that (44) can be calculated using the Fourier transform technique.

It can also be shown that \( \tilde{I}_B(\Delta_2, \cdots, \Delta_P) \) in (17) can be approximated by \( \tilde{I}_B(\Delta_2, \cdots, \Delta_P) \) [8]

\[
\tilde{I}_B(\Delta_2, \cdots, \Delta_P) = -\frac{1}{2} \int_{\Gamma_1} g_n(\mathbf{x}, \mathbf{u}; \Delta_2, \cdots, \Delta_P) d\mathbf{x}
\]

where \( g_n(\mathbf{x}, \mathbf{u}; \Delta_2, \cdots, \Delta_P) \) is the causal solution of the matrix integral equation in (40). The motion coefficients can be estimated by maximizing the log-likelihood function in (11). The maximization is carried out using the steepest ascent method. Define \( \Theta = [\theta_2 \ \theta_3 \ \cdots \ \theta_P] \), where \( \theta_k = [d_{k11} \ d_{k12} \ \cdots \ d_{k21} \ d_{k22} \ \cdots]^T, k = 2, 3, \cdots, P \). Let \( \hat{\theta}_k \) represent the estimates of motion parameters. The maximization is done iteratively as follows

\[
\hat{\theta}_k^+ = \hat{\theta}_k^- + \gamma_k \frac{\partial \ln \Lambda(\Theta)}{\partial \theta_k} \bigg|_{\theta = \hat{\theta}_k^-}, \quad k = 2, 3, \cdots, P,
\]

where \( \gamma_k = [\gamma_{k11} \ \gamma_{k12} \ \cdots \ \gamma_{k21} \ \gamma_{k22} \ \cdots]^T \) are the convergence parameters which control the speed of convergence, “+” and “−” represent the present and previous estimates, respectively, and

\[
\frac{\partial \ln \Lambda(\Theta)}{\partial \theta_k} \approx \frac{\partial \tilde{I}_R(\Theta)}{\partial \theta_k} + \frac{\partial \tilde{I}_B(\Theta)}{\partial \theta_k}.
\]

At each iteration, we find new estimates of the motion parameters and consequently, the LMMSE estimate of \( s(\mathbf{x}) \), i.e., \( \hat{s}_1(\mathbf{x}) \), and then use these values to update the log-likelihood function. Therefore, motion estimation and filtering are done simultaneously at each iteration. The flow chart in Fig. 2 demonstrates each step of the algorithm.

V. SIMULATION EXPERIMENTS

To verify the method presented in the previous sections, we study the case for which the motion vector field \( \mathbf{d}(\mathbf{x}) \) is represented by the affine transformation model represented below [9], [11]–[13]

\[
\mathbf{d}(\mathbf{x}) = (\mathbf{I}_2 - \mathbf{A})\mathbf{x} - \mathbf{b}
\]

where \( \mathbf{I}_2 \) is the 2 \times 2 identity matrix, \( \mathbf{A} = [a_{ij}], i, j = 1, 2 \), and \( \mathbf{b} = [b_1 \ b_2]^T \). Then the estimation of the image motion field becomes the estimation of the matrix \( \mathbf{A} \) and the vector \( \mathbf{b} \). The affine motion model is chosen because it covers the important motion scenarios such as translation, rotation, skew, or scaling.

To measure the performance of the filtering of the reference frame, a quantity of improvement in signal-to-noise ratio (ISNR) is defined in dB as

\[
\text{ISNR} = 10\log_{10} \left( \frac{\sum_{x_1} \sum_{x_2} [s(x) - r_1(x)]^2}{\sum_{x_1} \sum_{x_2} [s(x) - \hat{s}(x)]^2} \right)
\]
A normalized mean squared error (NMSE) function to measure the performance of the multiple frame motion estimation at each iteration:

\[
\text{NMSE}_k(i) = \frac{1}{NF} \sum_{x_1} \sum_{x_2} \left[ s(x - d_k(x)) - s(x - \hat{d}_k(i)(x)) \right]^2,
\]

where \( NF \) is a normalization factor to set \( \text{NMSE}(0) = 1 \), \( i \) is the index of iteration, and \( \hat{d}_k(x) \) is the estimate of \( d_k(x) \).

The simulations are performed for \( P = 3, 5 \), and \( 5 \) and signal-to-noise ratio (SNR) of 5 and 10 dB. We select the third frame as the reference frame when \( P = 3 \). The steady state NMSE's and ISNR for the reference frame are listed in Table I which indicate that the results of motion estimation are less accurate as the frame number \( P \) increases. The degradation in the motion estimation accuracy as \( P \) increases stems from the simultaneous nature of multiple frame motion estimation. It is noticed, however, that the overall performance of filtering in terms of ISNR is still improved.

Fig. 3 shows the noisy image sequence when SNR = 10 dB for the case for which \( P = 3 \). Fig. 4 depicts the filtered image sequence. The results shown in Fig. 4 indicate that the noise is clearly reduced. The NMSE curves for this case is plotted in Fig. 5. The motion parameters used in this experiment and their estimated values are listed in Table II.

To test the upper bound of the filtering performance of the proposed method, we used the true motion parameters to filter the reference frame in (39). The upper bounds of ISNR for each case we tested are listed in Table I. Fig. 6 shows the ISNR curves obtained using the true motion parameters and the motion parameters estimated from the proposed multiple frame motion estimation (MFME) algorithm for the case \( P = 3 \) and SNR = 10 dB. It is seen from Fig. 6 that the resultant ISNR closely converges to the ISNR corresponding to the ideal scenario for which the motion field is completely known.

**Table II**

<table>
<thead>
<tr>
<th>( P )</th>
<th>True value</th>
<th>Estimated value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2 )</td>
<td>0.1993</td>
<td>0.1986</td>
</tr>
<tr>
<td>( 3 )</td>
<td>0.1993</td>
<td>0.1993</td>
</tr>
<tr>
<td>( 5 )</td>
<td>0.1993</td>
<td>0.1993</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.19866</td>
<td>0.19830</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>-0.19866</td>
<td>-0.19866</td>
</tr>
<tr>
<td>( a_{13} )</td>
<td>0.19866</td>
<td>0.19866</td>
</tr>
<tr>
<td>( a_{14} )</td>
<td>-0.19866</td>
<td>-0.19866</td>
</tr>
<tr>
<td>( a_{21} )</td>
<td>0.10066</td>
<td>0.10076</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.10066</td>
<td>0.10066</td>
</tr>
<tr>
<td>( a_{23} )</td>
<td>0.00097</td>
<td>0.00097</td>
</tr>
<tr>
<td>( a_{24} )</td>
<td>0.01993</td>
<td>0.01993</td>
</tr>
<tr>
<td>( a_{25} )</td>
<td>0.00007</td>
<td>0.00007</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( b_{13} )</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>( b_{14} )</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>( b_{15} )</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>( b_{21} )</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>( b_{23} )</td>
<td>0</td>
<td>-2</td>
</tr>
</tbody>
</table>
VI. SUMMARY
This paper has presented an iterative method that simultaneously performs multiple-frame image motion estimation and filtering of image sequences. Traditional motion estimation methods only estimate the motion between two frames. The method presented in this paper can estimate the motions of several frames relative to the reference frame simultaneously in very noisy environments. The estimated multiple-frame motion can possibly be applied to image sequence compression. The noisy reference frame is also filtered during each iteration of motion estimation. Simulations using the affine motion model have validated our method. The simulation results indicate that, as the number of frame increases, we get better performance for filtering in terms of ISNR. Simulation results also indicate that the resultant ISNR closely converges to the ISNR corresponding to the known motion field. This is a severe test which clearly validates the performance of the method.

REFERENCES

Chich-Min Fan (M’99) received the B.S. degree in geophysics from National Central University, Chung-Li, Taiwan, R.O.C., in 1984, the M.S. degree in geophysics and the M.S. degree in electrical engineering, both from Michigan Technological University, Houghton, in 1989 and 1992, respectively, and the Ph.D. degree in electrical engineering from The Catholic University of America, Washington, DC, in 1998. He is currently with the U.S. Patent and Trademark Office, Crystal City, VA.

His research interests are in image motion estimation, image sequence filtering/restoration, digital communications, and seismic data processing.

Nader M. Namazi (S’84–M’85–M’92) received the B.S. degree in electrical engineering from the Iran College of Science and Technology, Tehran, Iran, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of Missouri, Rolla, in 1981 and 1985, respectively.

From 1985 to 1992, he was with the Michigan Technological University, Houghton, where he was promoted to Associate Professor. Since 1992, he has been with the Catholic University of America, Washington, DC, where he is currently an Associate Professor of electrical engineering and computer science. His research interests are in image motion detection and estimation, image sequence filtering/restoration, and digital communications. He was the Editor-in-Chief of the International Journal of Modeling and Simulation from 1994 to 1996.

He is the author of New Algorithms for variable Time Delay and Nonuniform Image Motion Estimation (Norwood, NJ: Ablex, 1994).

Dr. Namazi is a member of Eta Kappa Nu.