Nonstationary AR Modeling and Constrained Recursive Estimation of the Displacement Field

Serafim N. Efstratiadis, Member, IEEE, and Aggelos K. Katsaggelos, Member, IEEE

Abstract—In this paper, an approach for the constrained recursive estimation of the displacement vector field (DVF) in image sequences is presented. An estimate of the displacement vector at the working point is obtained by minimizing the linearized displaced frame difference based on a set of observations that belong to a causal neighborhood (mask). An expression for the variance of the linearization error (noise) is obtained. Because the estimation of the DVF is an ill-posed problem, the solution is constrained by considering an autoregressive (AR) model for the DVF. This AR model is first considered stationary, according to which the two components of the DVF are uncorrelated and each component is modeled by a 2-D discrete Markov sequence. A nonstationary AR model of the DVF is also considered by spatially adapting the model coefficients using a weighted LMS algorithm. Additional information about the solution is incorporated into the algorithm using a causal “oriented smoothness” constraint. Based on the above formulation, a set theoretic regularization approach is followed that results in a weighted constrained least-squares estimation of the DVF. The proposed algorithm shows an improved performance with respect to accuracy, robustness to occlusion, and smoothness of the estimated DVF when applied to typical videoconferencing scenes.

I. INTRODUCTION

In the area of image sequence processing, a very important task is the estimation of the velocity field, which results from the apparent motion of the image brightness, and which is also called optical flow [1]. That is, the apparent displacement of each picture element between successive frames forms is the so-called displacement vector field (DVF). The estimated DVF has been used in many applications, such as, motion-compensated (MC) coding, interpolation, and filtering (see, e.g., [2]–[6]).

A class of displacement estimation techniques is represented by pel-recursive (PR) gradient-based algorithms [5], [7]–[9]. PR algorithms can provide displacement vectors with subpixel accuracy without the need of spatial interpolation. Also, because of the recursive computability of the DVF, they are suitable for applications, such as predictive MC compression schemes that do not require the transmission of the estimated DVF or in recursive MC filtering algorithms. PR algorithms are derived based on the minimization of the linearized displaced frame difference (DFD), using a set of points that belong to a causal mask around the working point, under the assumption that the image intensity remains constant along the motion trajectories. The resulting linearization error term can be assumed to be a sample of a stochastic process (noise). Then, based on the assumption that the displacement vector is also a sample of a stochastic process, a class of Wiener-based PR algorithms has been proposed [7], [10], [11]. However, due to the stationarity assumption about the solution, Wiener-based algorithms are not robust to the discontinuities of the DVF. Toward that end, an improved adaptive multiple-input weighted least-squares algorithm has been proposed assuming that the displacement vector is a deterministic unknown-bounded signal [12], [13]. In addition, prior information about the solution is incorporated into the algorithm using a causal “oriented smoothness” constraint [13], that is, by minimizing the variation of the DVF along the direction perpendicular to significant image intensity variations [14].

In this paper a constrained gradient-based recursive algorithm for the estimation of the DVF is presented. A smoothness constraint is imposed on the estimated solution with the use of an autoregressive (AR) model for the DVF. Such a model is stationary or it adapts along the direction of recursion. Stochastic models for the DVF have been used previously in a different context. For example, the authors used in [11] a 3-DMC AR model for the DVF for obtaining an initial estimate required by the pel-recursive algorithm. Namazi and Lee [15] used the modeling of each component of the DVF by a 2-D discrete Markov sequence with a given exponential autocorrelation function. Then they proposed on iterative displacement estimation algorithm based on the generalized maximum likelihood criterion. Konrad [16] used a compound Markov vector model for the DVF and applied it to the restoration of the DVF. A Kalman estimator was
also developed in [17], for the restoration of the DVF. Recently, vector models similar to Konrad’s model have been proposed in [18], [19], which are applied to the task of estimating the DVF. Also, vector-coupled Gauss–Markov models have been proposed in [20], [21], which led to the development of recursive MAP estimators for the DVF.

More specifically, a set theoretic formulation is followed in this work by using an estimate of the variance of the linearization error (observation noise). This approach results in a weighted constrained least-squares estimation of the DVF where prior information about the solution is incorporated in the form of weights. According to the adaptive multiple-input approach [12], [13] that is followed, 1) regularization, namely, improvement of the condition of the matrix to be inverted, is provided directly, and 2) an estimator with spatially variable size mask results, due to the weighting of the observations according to the estimated variance of the linearization error at each pixel. In addition, an “oriented smoothness” constraint (OSC) [14] is adopted for weighting the prior knowledge about the solution that is represented by the AR model of the DVF. The use of OSC in our recursive formulation also results in a spatially adaptive causal prediction model for the DVF. Finally, the nonstationarity of the DVF is considered by adopting an adaptive least-mean-squares (LMS) algorithm. The AR model coefficients are now updated along the horizontal scanning direction. The proposed algorithm has a significantly improved performance compared with the Wiener-based algorithms of [7]. It also results in a smoother DVF than recursive algorithms without the modeling of the DVF [12], [13].

The paper is organized as follows. In Section II, the recursive DVF estimation using a multiple input solution approach and the AR modeling of the DVF are presented. In Section III-A, the proposed recursive DVF estimation method using a set theoretic formulation is derived. The determination of the various parameters and weight matrices is discussed in Section III-B, and an adaptive prediction model for the evaluation of the initial displacement estimate is presented in Section III-C. An adaptive weighted LMS algorithm for the nonstationary AR predictive modeling of the DVF is presented in Section IV. Various implementation issues are discussed and a number of experiments with typical video-conferencing scenes are shown in Section V. Finally, the conclusions drawn from our experiments are given in Section VI.

II. PROBLEM FORMULATION

A. Gradient-Based Recursive DVF Estimation

Let \( f_k(r) \) denote the image intensity at the spatio-temporal position \((r, k)\), where \( r = [x, y]^T \), and \( k \) is the frame number. In the current frame \( f_k(m, n), 1 \leq m \leq M, 1 \leq n \leq N \), let \( r \) represent the working point \((m, n)\). Given two successive frames \( f_k \) and \( f_{k-1} \), a DVF namely \( [d(r)] \), is defined as the 2-D vector field that maps the points in \( f_{k-1} \) onto their corresponding locations in \( f_k \). For simplicity, \( d = [d_x, d_y]^T \) denotes the displacement vector corresponding to the working point \( r \) in the current frame \( f_k \). The goal is to find an estimate \( \hat{d} \) based on values of \( f_k \) and \( f_{k-1} \) in a neighborhood of \( r \).

Let us assume that the working pixel belongs to a moving area. Then, the points around \( r \) can also be assumed to belong to the same area. More specifically, consider a mask \( \mathcal{M} \) that contains the working point and \( L - 1 \) additional points such that it allows for the recursive computation of \( \hat{d} \). Such a mask is shown in Fig. 1(a) for \( L = 5 \), with the points denoted by \( r_i \), \( i \in \mathcal{M} \), where \( \mathcal{M} = \{0, 1, \ldots, L - 1\} \) is the set of indexes of the points in the mask. Any mask \( \mathcal{M} \) can be decomposed into \( r \) sub-masks \( \mathcal{M}_i \), where \( \mathcal{M} = \bigcup_i \mathcal{M}_i \), and \( \bigcup \) denotes union of sets. Each submask \( \mathcal{M}_i \) contains the working point \( r_0 \) and has a total of \( L \) points. For example, for \( v = 2 \), we have the standard approach namely, \( \mathcal{M}_i = \mathcal{M} \) and \( L_i = L = 5 \). If we let \( v = 2 \) and \( L = 3 \), then an example of two submasks is shown in Fig. 1(b), that is, \( \mathcal{M}_1 = \{0, 1, 2\} \) and \( \mathcal{M}_2 = \{0, 3, 4\} \). Finally, \( v = 4, 5 \) systems of equations with \( L = 2 \) equations each can be formed based on four one-point overlapping submasks \( \mathcal{M}_i \), that is, \( \mathcal{M}_i = \{0, i\} \), for \( i = 1, 2, 3, 4 \) (Fig. 1(c)).

Let \( d(r_0) \) now be the displacement vector of the 0th point in \( \mathcal{M} \), and let \( \hat{d} \) be an initial estimate of \( d = d(r_0) \). By associating the same initial estimate \( \hat{d} \) with all points in \( \mathcal{M} \), the displacement update vector \( u(r_0) = [u(r_0)u(r_0)]^T \) is equal to

\[
    u(r_0) = d(r_0) - \hat{d} \tag{1}
\]

with \( u(r_0) = u \). In the derivation of displacement estimation algorithms, it is commonly assumed that the image intensity is constant along the motion trajectories. Based on this assumption, let us define the displaced frame difference (DFD) at \( r \) as follows:

\[
    \Delta_s(r, u(r_0)) = f_k(r) - f_{k-1}(\hat{r} + u(r_0)) = f_k(r) - f_{k-1}(r - \hat{d})
\]

\[
    = f_k(r) - f_{k-1}(\hat{r} + u(r_0)) \tag{2}
\]

where \( \hat{r} = r - d(r_0) \) is the (unknown) location of point \( r_0 \) in frame \( f_{k-1} \) along the motion trajectory. If we assume a uniform DVF inside the mask and a good available initial estimate \( \hat{d} \), then, each \( u(r_0) \) is fairly small. Then, by linearizing \( \Delta_s \) using the Taylor series expansion of \( f_{k-1}(r) \) at the location \((\hat{r} + u(r_0))\), we get

\[
    f_{k-1}(\hat{r} + u(r_0)) = f_{k-1}(\hat{r}) + \nabla f_{k-1}(\hat{r})(u(r_0)) + u(r_0) \tag{3}
\]

where \( \nabla = \left[ \frac{\partial}{\partial x} \frac{\partial}{\partial y} \right]^T \) represents the spatial gradient operator and \( u(r_0) \) is the linearization error term. Using (2) and (3), we further obtain

\[
    \Delta_s(r, u(r_0)) = -g_k(\hat{r}, u(r_0)) = -g_k(\hat{r}, u(r_0)) \tag{4}
\]

where \( g_k(\hat{r}, u(r_0)) \) is the linearization error term. Using (2) and (3), we further obtain

\[
    \Delta_s(r, u(r_0)) = -g_k(\hat{r}, u(r_0)) \tag{4}
\]
where
\[
\begin{bmatrix}
g_{x}(l)g_{y}(l)
\end{bmatrix} = \nabla^T \phi_{k-1}(r_j - \hat{d}).
\] (5)

An estimate of the gradient vector is computed numerically as discussed in Section V. Equation (4) expresses the basic equation for estimating \( \hat{d} \) based on \( \hat{d} \) and knowledge of the DFD and the spatial gradients. The linearization error (or noise) term \( v_{k}(r_j, u(r_j)) \) plays an important role in the estimation and it is considered to be a sample of a stochastic process \( [7], [10], [13], [22] \). The variance of \( v_{k}(r_j, u(r_j)) \) depends on the spatial location \( r_j \) and is small or large depending on whether the DVF at the area around \( r_j \) is uniform or discontinuous, respectively (see Section III-B-1).

In obtaining a workable expression for (4), we need to consider only one unknown update vector. Thus, in (4), 1) we set \( u(r_j) = \hat{u} \) for the term in the inner product with the gradient vector and 2) we keep the dependence of \( u \) on \( r_j \) for the error term \( v_{k}(r_j, u(r_j)) \). Assumption 1) is valid for small mask sizes due to the smoothness of the DVF. However, at the discontinuities of the DVF, the variability of the DVF within the mask is significant. Thus, the error due to assumption 1) is taken into account by using 2), namely, by considering \( v_{k}(r_j, u(r_j)) \) as a function of \( u(r_j) \) at each point inside the mask. Note that because estimates of the displacement vector for the points in \( M \) other than \( r_j \) are available, an estimate of \( u(r_j) \) is also available and is considered in obtaining an estimate of the variance of \( v_{k}(r_j, u(r_j)) \) (Section III-B-1). Since \( \Delta_k \) is scalar, there may be more than one \( u \) vector that satisfy \( \Delta_k(r_j, u(r_j)) = 0 \). Therefore, a unique solution can be obtained by considering at least two observation points, provided that the estimation of the DVF at location \( r \) is not prohibited by the aperture effect [23]. In other words, there exist locations in the image sequence where the component of the optical flow along the isobrightness contours can not be determined from the intensity field. In general, however, we assume that an estimate \( \hat{u} \) can be obtained by solving a suitable system of equations based on (4).

Based on assumption 1), the \( L \) equations in (4) are combined into
\[
z_i = G_i \hat{u} + v_i \quad \text{for} \quad i = 1, 2, \cdots, \nu
\] (6)

where the matrix \( G_i \) represents an \( L \times 2 \) known deterministic gradient operator and \( z_i \) and \( v_i \) represent \( L \times 1 \) vectors resulting from stacking \( \Delta_k(r_j, u(r_j)) \) and \( v_{k}(r_j, u(r_j)) \), respectively. Then the displacement estimation problem is to find a displacement update \( \hat{u} \) as close as possible to the original one, subject to a suitable optimality criterion, given the observations \( z_i \), for \( i = 1, 2, \cdots, \nu \). In this paper, (6) is solved by following a set theoretic regularization approach [24]–[26]. The linearization error is taken into account by assigning weights to the various observations \( z_i \), and prior knowledge about the solution is used by considering a spatial AR model for the DVF [27].

B. Stationary AR Modeling of the DVF

The DVF is characterized in general by a high degree of spatial correlation. Let us assume that \( d(r) = [d_1(r), d_2(r)]^T \) is a random vector whose components \( d_i \) are \( 2 \)-D random variables \( [28] \). By following a stochastic formulation, the true DVF is considered to be a stationary \( 2 \)-D random field, with autocorrelation function the following symmetric positive definite matrix
\[
R_d(s) = E[d(r)d(r-s)^T] =
\begin{bmatrix}
R_{11}(s) & R_{12}(s) \\
R_{21}(s) & R_{22}(s)
\end{bmatrix}
\] (7)

where \( s = (s_1, s_2) \), \( R_{ij}(s) \) is the cross-correlation function between components \( d_i \) and \( d_j \), and \( E[] \) denotes expectation. Based on the above model, the prediction equation can be written as
\[
d(r) = \sum_{s \in S_d} A(s)d(r-s) + e(r)
\]
\[
= \hat{d}(r) + e(r)
\] (8)

where \( \hat{d}(r) \) is the predicted displacement vector, \( e(r) \) is the prediction error vector,
\[
A(s) = \begin{bmatrix}
a_{11}(s) & a_{12}(s) \\
a_{21}(s) & a_{22}(s)
\end{bmatrix} \quad \forall s \in S_d
\] (9)

are the coefficient matrices of the AR model, and \( S_d \) is the causal spatial support. In (9), matrices \( A(s) \) can be evaluated by solving the normal equations, that is,
\[
R_d(r) - \sum_{s \in S_d} R_d(r-s)A(s)^T = 0, \quad \forall r \in S_d.
\] (10)

The autocorrelation model of (7) is general and can describe a large number of stochastic vector signals. However, this model becomes computationally very tractable by assuming that the vertical and horizontal displacement components are each modeled as discrete \( 2 \)-D Markov sequences [3]. More specifically, by considering an expon-
nential model for each component and assuming that the
two components are uncorrelated, the following general
form is used
\begin{equation}
R_i(s) = \begin{cases} 
\sigma_i^2 \rho_i^{|i-j|} \sigma_j^2 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases} \quad (11)
\end{equation}

where \( \sigma_i^2 \) is the variance of \( d_i(r) \), and \( \rho_i \) and \( \rho_j \) are its
 correlation coefficients in the vertical and horizontal
directions, respectively, where \( 0 < \rho_i < 1 \). In this case, \( J_i \)
represents a first-order spatial quarter-plane support, that is, \( J_i = \{(1, 0), (0, 1), (1, 1)\} \). By using the exponential
model of (11) into (10), we obtain the following prediction
coefficients:
\begin{equation}
a_{ij}(s) = \begin{cases} 
\left( -\rho_i \right)^{|i-j|} \left( -\rho_j \right)^{|i-j|} & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases} \quad \forall s \in J_i. 
\end{equation}

In evaluating \( a_{ij}(s) \), the correlation coefficients \( \rho_i \) are
estimated from the available estimated DVF in the previous
frame \( \{d_{i-1}\} \), under the assumption that there is no
significant change in the optical flow between successive
frames (e.g., a scene change). Also, the theoretical expres-
sion of the autocorrelation matrix of the prediction error
vector \( e(r) \) is given by
\( R_i(s) = \text{diag}(\sigma_i^2, \sigma_i^2) \), where
\( \sigma_i^2 = (1 - \rho_i^2)(1 - \rho_j^2) \sigma_i^2. \quad (13) \)

Based on the above modeling of the DVF, the pre-
predicted displacement vector \( \hat{d} \) represents prior
knowledge about the solution at the working point and needs to be
considered in the estimation of \( \hat{d} \), as discussed in Section
III. However, due to the nonstationarity of the true DVF, the
AR model needs to be spatially adaptive in order to
model more accurately the variations of the DVF at
different moving areas in real video sequences. Toward
that end, a recursive adaptive LMS algorithm is presented
in Section IV.

III. CONstrained RECURSIVE DISPLACEMENT
ESTIMATION

A. SET THEORETIC FORMULATION

In estimating a solution to (6) a set theoretic approach
is followed [24], [25]. More specifically, let the lineariza-
tion error process \( v_i = z_i - G_i u \), which represents the
uncertainty in the observation \( z_i \), belong to a set
\( Q_i = \{v_i : \|v_i\|_{\hat{w}_i} \leq \epsilon_i^2\} \). Then, \( u \) belongs to a new set
\begin{equation}
Q_{u/z_i} = \left\{u : \|z_i - G_i u\|_{\hat{w}_i}^2 \leq \epsilon_i^2\right\} 
= \left\{u : \|u - u^*_i\|^2 \Omega_i [u - u^*_i] \leq 1\right\} 
\text{for } i = 1, 2, \ldots, \nu \quad (14)
\end{equation}

where \( \Omega_i = (1/\epsilon_i^2) G_i^T W G_i \), \( u^*_i = G_i^T z_i \), is the
center of the \( i \)th ellipse, \( G_i^T \) is the generalized inverse of
\( G_i \), and \( \|X\|_{\hat{w}} = \sum_{i=1}^{\nu} \epsilon_i^2 \sum_{i=1}^{\nu} \|X\|_{\hat{w}_i} \) represents a weighted norm. \( \Omega_i \) is a \( 2 \times 2 \)
positive definite matrix whose eigenvectors and eigenval-
does determine the orientation and the lengths of the
 corresponding axes of \( Q_{u/z_i} \). Similarly, prior knowledge
about \( u \) is used in the form of constraints. That is, given a
vector \( a, e = u - a \), which represents the uncertainty in
restricting the solution, is assumed to belong to set \( Q_a \).
If we set \( Q_a = \{e : \|e\|_{\hat{w}_a} \leq E^2\} \), then \( u \) is restricted to lie in the set
\begin{equation}
Q_{a/u} = \left\{u : \|u - a\|^2 \leq E^2\right\} = \left\{u : \|u - a\|^2 \Omega [u - a] \leq 1\right\} 
\end{equation}

where \( \Lambda = (1/E^2) F_a \) and \( a \) is the center of the ellipse.
Here, \( \Lambda \) (or \( F_a \)) is a \( 2 \times 2 \) positive definite matrix that
determines \( Q_{a/u} \). The choice of \( \epsilon_i, W_i, a, F_i, \) and \( E \)
is discussed in Section III-B.

Consider now sets \( Q_{a/u} \) and \( Q_{a/z_i} \). Each set contains \( u \)
and, therefore, \( u \) must lie in their intersection, denoted by
\( Q_a \). That is,
\begin{equation}
Q_a = \left( \bigcap_{i=1}^{\nu} Q_{a/z_i} \right) \cap Q_{a/u} \quad (16)
\end{equation}

where \( \bigcap \) denotes set intersection. According to (16),
the estimated set is the smallest one that must contain \( u \)
and that can be calculated from the available information [24].
One approach to geometrically describe the intersection
\( Q_a \), which is not necessarily an ellipse, is to consider an
ellipse that bounds it [24]. Then, the center of this ellipse
is chosen to represent the estimated solution vector. Fig. 2
shows the case of three intersecting ellipses, namely,
\( Q_{a/z_1}, Q_{a/z_2}, \) and \( Q_{a/z_3}, \) their intersection \( Q_a \) (shaded
area), and the center \( \hat{u} \) of the bounding ellipse. Given
the ellipses of (14) and (15), the center of a bounding
ellipse is described by [25]
\begin{equation}
\hat{u} = \sum_{i=1}^{\nu} \frac{1}{\epsilon_i^2} G_i^T W G_i + \frac{1}{E^2} F_a \nu = \sum_{i=1}^{\nu} \frac{1}{\epsilon_i^2} G_i^T W G_i + \frac{1}{E^2} F_a a. 
\end{equation}

Note that the contribution of each observation and con-
straint to the solution is inversely proportional to the size
of the corresponding bound, namely, \( \epsilon_i^2 \) (amount of noise
present in the \( i \)th input) and \( E^2 \) (uncertainty in the
constraint). Equation (17) can also be derived by combining
quadratically the constraints expressed by (14) and
(15), and then minimizing the functional
\begin{equation}
\Theta(u) = \sum_{i=1}^{\nu} \frac{1}{\epsilon_i^2} \|z_i - G_i u\|_{\hat{w}_i}^2 + \frac{1}{E^2} \|u - a\|^2 
\Xi(u) + \Psi(u) \quad (18)
\end{equation}

with respect to \( u \). The necessary condition for \( \Theta(u) \) to
have a minimum is that its gradient with respect to \( u \) be
equal to zero. It can be easily shown that this condition
leads to (17) [25]. Such an approach, however, cannot be
fully explained unless it is considered in the framework of
the set theoretic regularization, as is done here. Such a
framework is further essential in choosing the weight
matrices, as it becomes apparent in the later sections of
the paper.
Let us now consider various special cases that result from the above formulation. For example, when $E$ becomes infinite, the set $Q_{a, e}$ does not constrain the solution any more. As another case, if we assume that one of the bounds $e_i$ or $E$ is not available, then the estimation problem is formulated as the minimization of the size of the corresponding ellipse subject to the requirement that the solution belongs on the surface of the remaining ellipses. For example, the constrained least-squares solutions $\hat{u}_i$, $\hat{u}_a$, and $\hat{u}_{\epsilon i}$ are shown in Fig. 2, for the respective cases when the bounds $E$, $e_i$, and $e_{\epsilon i}$ are not defined. Therefore, when both $e_i$ are defined, then $\hat{u}_a$ is closer to the centers of the two ellipses, which represent extremely rough solutions. Otherwise, when $E$ is defined the solution moves closer to the center of the ellipse $Q_{a, e}$, which represents an extremely smooth solution. Thus, the estimate $\hat{u}_a$ represents a trade-off between extremely rough and extremely smooth solutions. In applying the above formulation to the DVF estimation problem, matrices $W_i$, $F_i$, as well as vector $a$ and parameters $e_i$ and $E$, are determined next.

B. Weight Matrices and Parameters

Choice of $W_i$ and $e_i$. Matrix $W_i$ in (14) contains prior knowledge about the importance of each linearization error term $v_i_i(r_i, u(r_i))$, where $i \in A$. More specifically, $v_i_i(r_i, u(r_i))$ for each mask $A_i$ represents a sample of a zero-mean stochastic process with covariance matrix

$$R_i = E[v_i v_i^t] = diag[c_i(l), l \in A]$$

for $i = 1, 2, \ldots, \nu$. (19)

An estimate of the variance $c_i(l)$ is given by [12] (see also the Appendix)

$$\hat{c}_i(l) = \frac{1}{4} \sum_{(i,j)} \sigma_{ij}^2 \hat{u}_i(r_i)^2 \hat{u}_j(r_j)^2 + \sigma_i^2$$

(20)

where $\sigma_{ij}^2$ are the variances of the elements of the matrix $B_i = \nabla \nabla_i - (r_i - d_i)^2$, $\hat{u}_i(r_i) = d(r_i) - d_i$, and $\sigma_i^2$ is an estimate of the variance of the higher order linearization terms. Note that in previous work [7], [10], all $\hat{c}_i(l)'s$ were assumed to be equal, whereas in [29] the variance of the update vector was computed by assuming the same $\hat{u}$ for all points in $A_i$ ($\delta = 1$ was used in all cases). Since the eigenvalues of $W_i$ divided by $e_i^2$ are proportional to the length of the corresponding axes of $Q_{a, e}$, we set $W_i = \hat{c}_i(0) R_i^{-1}$, that is,

$$W_i = \begin{bmatrix} \frac{\hat{c}_i(0)}{\hat{c}_i(l)} \end{bmatrix}, \quad l \in A_i$$

for $i = 1, 2, \ldots, \nu$. (21)

Matrix $W_i$ needs to be evaluated based on prior information about the estimated DVF. More specifically, for the working point we have $\hat{d}(r_i) = d_i$, thus, $\hat{u}(r_i) = 0$, which results in $\hat{c}_i(0) = \min \{\hat{c}_i(l), l \in A_i \} = \sigma_i^2$ according to (20). That is, $w(0) = 1$, and $0 < w(l) \leq 1$, for $l \neq 0$, from (21). According to this equation, the weights $w(l)$ are inversely proportional to a measure of the variation of the DVF in $A_i$. For example, if we assume that the working point belongs to a moving area, whereas, the $l$th mask point belongs to a stationary area, then, $||u(r_i)|| \gg ||u(r_i)||$ and $w(l) \ll 1$. Therefore, the advantage of this weighted estimation approach is that, by computing the “degree of trust” of each observation, although the analysis window has fixed size, the overall effect is that of a variable size window. As a result, large errors in estimating the DVF that usually occur in occlusion areas are avoided.

A bound $e_i$ on the weighted norm of $v_i$ is determined by considering the weight matrix $W_i$ given by (21), that is,

$$E[||v_i||_{W_i}] = E[w(l) ||v_i||_{W_i}] = L_i \hat{c}_i(0) = \epsilon_i^2$$

where it is assumed that each of the estimated variances $\hat{c}_i(l)$ is equal to the true variance of the linearization error at $r_i$.

Choice of $F_i$, $a$, and $E$. Matrix $F_i$ in (15) should be chosen to describe known properties of the true DVF and its introduction should result in a linear system of equations (17), which is better conditioned than when $F_i = 0$ [25], [26]. An analytic expression for $F_i$ is obtained by considering a general smoothness constraint for the DVF, which is expressed by the requirement that the functional

$$\Phi(d) = \frac{1}{E_d^2} \text{trace} \left( \nabla^2 d \right) F_i \left( \nabla^2 d \right)$$

(23)

is minimized, where $E_d^2$ is a suitable normalization constant, $\nabla d = [\nabla d_1, \nabla d_2]$ is the spatial gradient of the vector $d$, and $F_i$ is a $2 \times 2$ weight matrix. The above smoothness constraint with $F_i = 1$ was first proposed by Horn [14]. It was later refined by Nagel and Enkelmann [16], who introduced the concept of the “oriented smoothness” constraint (OSC). According to this constraint, the variation of $d$ in the direction perpendicular to both the gradient $\nabla d$ and the principal curvature direction (eigen-vector of $\nabla^2 d$) associated with a large curvature (eigen-value of $\nabla^2 d$) at the working point should be as small as possible [14], [30]. In other words, the variation of $d$ is minimized in the direction perpendicular to significant image intensity variations. Implementing the minimization based on the above definition results in a weight matrix
with structure analogous to that of matrix \( \tilde{\Omega}^{-1} \), where 
\( \tilde{\Omega} = \tilde{G}^2 \tilde{G} \) and \( \tilde{G} \) is formed using entries \([\tilde{g}_i, (1) \tilde{g}_j, (1)] = \nabla f_\lambda (r) \) over a square noncausal mask (window) centered at the working point [14]. Therefore, we choose

\[
F_d = \tilde{\Omega}^{-1}
\]

where \( \tilde{\Omega} = \sum_{i=1}^{L} \tilde{\Omega}_i \) and each \( \tilde{\Omega}_i \) is formed by the pixels in the \( i \)th submask. Note that \( \tilde{\Omega} \) does not depend on the DVF but only on the intensity values of the current frame, whereas \( \Omega = \sum_{i=1}^{L} \Omega_i \) and each \( \Omega_i \) depends on an initial estimate \( \tilde{d} \), which is used in the estimation of \( u \). In the case when \( \tilde{d} \) is a true displacement, the two matrices are identical, since they are both formed using intensity values that lie on the same motion trajectory, where the image intensity is assumed constant. Although each \( \tilde{\Omega}_i \) is usually ill conditioned, \( \tilde{\Omega} \) has significantly improved condition number [13].

Based on the above, let us assume that \( \tilde{\Omega} = \tilde{\Omega} \). This assumption is true for \( u = 0 \) and is approximately true for \( ||u|| \) very small, as is usually the case in practice. Then, matrices \( \tilde{\Omega} \) and \( \Lambda \) commute, which means that they share the same complete set of eigenvectors [31]. Let \( \omega_1^2 \) and \( \lambda_i^2 = (1/E^2) \alpha_i^2 \), for \( i = 1, 2 \), denote the eigenvalues of \( \Omega \) and \( \Lambda \), respectively, and \( \mathcal{P}(\Omega) = ||\Omega||_2 : ||\Omega||_2 = \omega_1^2/\omega_2^2 \) the condition number of \( \Omega \). Then, since \( \omega_1 \geq \omega_2 \) and \( \lambda_1^2 \leq \lambda_2^2 \), inequality

\[
\mathcal{P}(\Omega + \Lambda) \leq \mathcal{P}(\Omega)
\]

is satisfied for any value of the parameter \( E^2 > 0 \). Inequality (25) indicates that \( \Lambda \) is such that, \( \lambda_1^2 = 0 \), therefore, the large eigenvalue of \( \Omega \) is not altered, while \( \lambda_1^2 \) is proportional to \( \mathcal{P}(\Omega) \) so that the smaller eigenvalue of \( \Omega \) is increased by an amount proportional to the degree of ill conditionedness of \( \Omega \). A related approach was proposed in [10], [11] for the single-input case, where the choice \( \Lambda = (\omega_1^2/\omega_2^2)I \) was shown to give improved estimation results.

In our recursive formulation, the variation of \( d \) in (23) is expressed by the innovation term of the AR model and a suitable normalization constant is the bound \( E^2 \), as defined in (15). Therefore, by considering a recursive OSC, the minimization of the following functional

\[
\Psi(d) = \frac{1}{E^2} \|e\|_{\tilde{F}_d}^2 = \frac{1}{E^2} \|d - \tilde{d}\|_{\tilde{F}_d}^2
\]

is used, which when written as a function of \( u \) results in the form given by (18), where

\[
a = \tilde{d}(r) - \tilde{d}
\]

and \( \tilde{d} \) is computed based on the available estimates. In order to determine the bound \( E \), we let assume that the uncertainty in restricting the solution or the AR model prediction error, \( e \), is a zero mean white stochastic process. That is, by letting \( E[ee'] = \text{diag}[\sigma_1^2, \sigma_2^2] \), it is easily shown that

\[
E\|u - a\|_{\tilde{F}_d}^2 = E\|e\|_{\tilde{F}_d}^2 = \xi_{11} \sigma_1^2 + \xi_{22} \sigma_2^2 = E^2
\]

where \( \xi_{ij} \) are the entries of matrix \( F_d \). Values of \( \sigma_1^2, \sigma_2^2 \), \( i = 1, 2 \), can either be computed by (13) or from the estimated DVF, that is, \( \sigma_1^2 = E(\tilde{d} - d)^2 \). Note that, if \( \sigma_1^2 = \sigma_2^2 \), then the term \( E(\tilde{d} - d)^2 \) in (28) represents a suitable normalization term for the matrix \( F_d \), as was also observed in [14]. Finally, the initial estimate \( \tilde{d} \) is computed based on the AR model and the OSC, as discussed in the next section.

C. Spatially Adaptive DVF Prediction

The initial estimate \( \tilde{d} \) is very important since it determines \( \tilde{a} \), the magnitude of vector \( \tilde{u} \), matrix \( \tilde{\Omega} \), and its condition, as well as the variance of the linearization error. Consequently, \( \tilde{d} \) determines the accuracy in estimating the DVF from (17). An appropriate initial estimate for \( d \) is obtained by using the prediction model of (8) and OSC. The resulting prediction model should be spatially adaptive in order to take into account the discontinuities of the DVF, which occur primarily around the boundaries of moving areas. Since matrix \( F_d \) in (24) contains information about the local spatial intensity variations at the working point, it is used to provide information about the discontinuities of the DVF.

More specifically, let us consider a vertical and a horizontal 1-D prediction model for the DVF with supports \( J = \{1, 0\} \) and \( J = \{0, 1\} \), respectively, and the corresponding model coefficient matrices \( A_j(s) \), where \( s \in J \), and \( \tilde{J} \), \( i = 1, 2 \). The elements of matrices \( A_j \), and \( A_{\tilde{J}} \) are given by (12) with the appropriate region of support \( J \). Then, the corresponding prediction errors of these models are

\[
e_i(r) = d(r) - A_j(s)d(r - s) \forall s \in J_i.
\]

It can be easily shown that the minimization of the functional

\[
T(d) = \text{trace} \left[ \{e_1, e_2\}^T F_d \{e_1, e_2\} \right]
\]

results in the following prediction equation

\[
\tilde{d} = \gamma(1, 0) A_j(1, 0) \tilde{d}_{10} + \gamma(0, 1) A_{\tilde{J}}(0, 1) \tilde{d}_{01}
\]

where \( \tilde{d}_{10} = \tilde{d}(r - (0, 0)), \tilde{d}_{01} = \tilde{d}(r - (0, 1)) \), and

\[
\gamma(1, 0) = \frac{1}{2} \left( 2 \xi_{11} + \xi_{12} + \xi_{21} + \xi_{22} \right)
\]

\[
\gamma(0, 1) = \frac{1}{2} \left( \xi_{11} + \xi_{12} + 2 \xi_{21} + \xi_{22} \right).
\]

Note that since

\[
\gamma(1, 0) + \gamma(0, 1) = 1
\]

(31) represents a causal prediction model for the DVF, which is spatially adapted based on information about the local intensity gradients. Special cases of (31) are the simpler approaches for evaluating \( \tilde{d} \), which have been described in [12], [13] with \( A_j(s) = I \). For example, by setting \( \gamma(1, 0) = 0 \), the simple horizontal prediction \( \tilde{d} = d_{10} \) was used in [7], [8], or for \( F_\tilde{J} = I \) we get \( \gamma(1, 0) = \gamma(0, 1) = 1/2 \), which is the simple 2-D prediction model used in [10], [22].
In order to demonstrate the performance of the above prediction model, let us assume that at each working point \( \mathbf{r} \) the local coordinate system is aligned with the eigenvectors of \( \hat{\Omega} \). Then, we can write that \( \hat{\Omega} = \text{diag}[\hat{\omega}_1^2, \hat{\omega}_2^2] \), where \( \hat{\omega}_1^2 \geq \hat{\omega}_2^2 \) are the eigenvalues of \( \hat{\Omega} \), and \( F_\delta = \text{diag}[\xi_{11}, \xi_{22}] = \text{diag}[\hat{\omega}_1^{-2}, \hat{\omega}_2^{-2}] \), with \( \xi_{12} = 0 \). Fig. 3 illustrates a typical displacement estimation problem of a uniformly moving rigid object, denoted by \( \mathscr{B} \), on a stationary background, denoted by \( \mathscr{A} \). For simplicity, the intensity values \( f_\mathscr{A} \) and \( f_\mathscr{B} \) at both areas \( \mathscr{A} \) and \( \mathscr{B} \) of the image are assumed constant and \( f_\mathscr{B} = f_\mathscr{A} \). Five displacement prediction cases are considered, one at an arbitrary point \( \mathbf{r} = \mathbf{r}_a \) inside the object, and four other at points \( \mathbf{r} = [r_1, \ldots, r_s] \) around the boundaries of the object. In each case, \( \mathscr{X} \) denotes the working point \( (m, n) \) and \( \mathscr{X}_a \) the prediction support points \( \{(m-1, n), (m, n-1)\} \).

For \( \mathbf{r} = \mathbf{r}_a \), it is clear that, for any matrix \( F_{\delta a} \), (31) provides a correct choice for \( \hat{\delta}_a \). For \( \mathbf{r} = [r_1, \ldots, r_s] \), the corresponding ellipses that are formed by the eigenvalues and eigenvectors of \( \hat{\Omega} \) are plotted in Fig. 3. At each of these points we have that \( \hat{\Omega}(\mathbf{r}) \gg 1 \) and the eigenvector associated with the large eigenvalue is directed perpendicular to the edge, whereas the eigenvector associated with the small eigenvalue is directed parallel to the edge [14], [30]. Thus, at \( \mathbf{r}_a \) we have \( \hat{\omega}_1^2 \gg \hat{\omega}_2^2 \) or \( \xi_{11} \gg \xi_{22} \), which gives \( \gamma(1, 0) = 1 \), \( \gamma(0, 1) = 0 \) and at \( \mathbf{r}_a \) we have \( \hat{\omega}_1^2 \ll \hat{\omega}_2^2 \) or \( \xi_{11} \ll \xi_{22} \), which gives \( \gamma(1, 0) = 0 \), \( \gamma(0, 1) = 1 \). Therefore, at \( \mathbf{r}_a \) we have \( \hat{\delta}_a = A_1(0, 1) \mathbf{d}_1 \) whereas at \( \mathbf{r}_a \) we have \( \hat{\delta}_a = A_1(0, 1) \mathbf{d}_1 \). Note that in Fig. 3 the suppressed points of the displacement prediction support at locations \( \mathbf{r}_a \) and \( \mathbf{r}_b \), namely, points \( (m-1, n) \) and \( (m-1, n) \), respectively, are denoted by "\( \circ \)". Finally, at \( \mathbf{r}_a \) and \( \mathbf{r}_b \), both prediction points are suppressed since both support points belong to different moving areas than the working point. At these points, the prediction model fails to give a good initial estimate since \( \hat{\Omega}(\mathbf{r}) \gg 1 \) and \( \gamma(1, 0) = \gamma(0, 1) = 1 \), therefore, \( \mathbf{d}(\mathbf{r}) = 0 \).

In conclusion, the above displacement prediction algorithm which is based on matrix \( F_{\delta a} \) and (31) allows the correct evaluation of \( \hat{\delta}_a \) at the discontinuities of the DVF, where other simpler prediction methods fail. For example, if we consider location \( \mathbf{r}_a \), then, the horizontal prediction approach used in [7], [8] gives \( \hat{\delta}_a = 0 \), whereas the approach proposed in [10], [22], which results from (31) by setting \( A_1(0, 1) = A_2(0, 1) = F_{\delta a} = 1 \), gives \( \hat{\delta}_a = \frac{1}{2} \mathbf{d}_1 \).

IV. ADAPTIVE NONSTATIONARY AR MODELING OF THE DISPLACEMENT FIELD

In this section an adaptive 2-D AR predictor for the DVF is presented. It uses the spatial correlation of the DVF in a small neighborhood around the working pixel and the prediction error at those points in order to adapt the model coefficients of the AR model. The proposed method is an application of a weighted adaptive least-mean-squares (LMS) filter [32] to the nonstationary AR modeling of the DVF. The algorithm updates the AR coefficients by using the available estimated DVF and the mean-squared-error (MSE), which is weighted appropriately by matrix \( F_{\delta a} \) that brings in prior information about the DVF from the intensity field of the working frame.

Let us consider the spatially varying form of (8), that is,

\[
d(\mathbf{r}) = \sum_{s \in \mathcal{S}_d} A(s, \mathbf{r}) d(\mathbf{r} - s) + e(\mathbf{r})
\]

(34)

where

\[
A(s, \mathbf{r}) = \begin{bmatrix} a_{11}(s, \mathbf{r}) & a_{12}(s, \mathbf{r}) \\ a_{21}(s, \mathbf{r}) & a_{22}(s, \mathbf{r}) \end{bmatrix} \quad \forall s \in \mathcal{S}_d
\]

(35)

and \( \mathcal{S}_d \) is an S-point causal spatial support. Let us consider matrix \( \tilde{\mathcal{A}}(\mathbf{r}) \) of size \( 2 \times 2S \), which is formed by stacking the S coefficient matrices \( A(s, \mathbf{r}) \), that is,

\[
\tilde{\mathcal{A}}(\mathbf{r}) = \begin{bmatrix} A(s, \mathbf{r})^T \\ \mathcal{S}_d^T \end{bmatrix}
\]

(36)

and vector \( \tilde{d}(\mathbf{r}) \) of size \( 2S \times 1 \) which is formed by stacking 5 displacement vectors, that is,

\[
\tilde{d}(\mathbf{r}) = [d(\mathbf{r} - s)]^T, \quad \forall s \in \mathcal{S}_d.
\]

(37)

Then, (34) is written as

\[
d(\mathbf{r}) = \tilde{\mathcal{A}}(\mathbf{r}) \tilde{d}(\mathbf{r}) + e(\mathbf{r}).
\]

(38)

The support that was proposed in Section II-B for the stationary AR model with \( S = 3 \) is also used in the adaptive filter. The adaptation of the above model is performed by applying the LMS algorithm in updating matrix \( \tilde{\mathcal{A}}(\mathbf{r}) \) along the direction of a horizontal image line. That is,

\[
\tilde{\mathcal{A}}(\mathbf{r} + \mathbf{k}) = \tilde{\mathcal{A}}(\mathbf{r}) - \beta \cdot \tilde{H}(\mathbf{r}) \quad \text{for} \quad k = (0, 1)
\]

(39)

where \( \beta \) is the adaptation gain, and \( \tilde{H}(\mathbf{r}) \) is a \( 2 \times 2S \) matrix of the weighted error function gradient. Similarly to matrix \( \tilde{\mathcal{A}}(\mathbf{r}) \), matrix \( \tilde{H}(\mathbf{r}) \) is defined by

\[
\tilde{H}(\mathbf{r}) = \nabla E[\|e(\mathbf{r})\|^2] = \begin{bmatrix} H(s, \mathbf{r})^T, \forall s \in \mathcal{S}_d \end{bmatrix}^T
\]

(40)
where

\[
H(s, r) = \frac{\partial}{\partial A(s, r)} E[\|e(r)\|_2^2]
= E \frac{\partial}{\partial A(s, r)} e(r)^T F_d e(r)
\quad \forall s \in \mathcal{S}_d
\tag{41}
\]

is the $2 \times 2$ gradient matrix of the weighted prediction error at the corresponding support point $s$. By using (34) and (35), it can be shown that

\[
\frac{\partial}{\partial A(s, r)} e(r)^T F_d e(r) = -2 F_d e(r) d(r - s)^T
\tag{42}
\]

resulting in

\[
H(s, r) = E[F_d e(r) d(r - s)^T].
\tag{43}
\]

Parameter $\beta$ determines the speed of the adaptation process; it is restricted in the range [32]

\[
0 < \beta < \frac{1}{SP_d}
\tag{44}
\]

where $P_d = E[d(r)^T F_d d(r)]$ is the input power of the DVF to the weighted filter and can be computed based on the estimated DVF. If we assume that the components of $d(r)$ are uncorrelated, then, we can set $P_d = \text{trace}(F_d^2 \sigma_d^2 + \sigma_d^2)$. In general, $\beta$ is determined by the amount of activity in the scene and needs to be adjusted.

In implementing this adaptation mechanism, a local estimate of the error function gradient for some interval in the past needs to be computed. As an example, the expectation in (43) can be estimated by

\[
H(s, r) = \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \phi(k) F_d e(r - k) d(r - k - s)^T
\tag{45}
\]

where $k = (k_1, k_2)$ and $\phi(k)$ represent the normalized weight coefficients or the window function, such that $\sum_k \phi(k) = 1$. Note that normalization is required in order for the adaptation gain $\beta$ to be independent of the window function and, therefore, allow a simplified tuning of the algorithm. Although the above averaging is expected to produce most reliable estimates when done in both directions with large values of $K_1$ and $K_2$, this is expensive computationally and storage wise. In our application, since a horizontal direction of recursion is used, the averaging is performed only in this direction, that is, by setting $K_1 = 1$ in (45). Several factors need now be considered when selecting a suitable value for $K_2$. A large value of $K_2$ results in slow adaptation of the coefficients, with no valuable results especially at the discontinuities of the DVF. On the other hand, a small value of $K_2$ results in faster convergence but more noisy estimates.

The determination of the weight coefficients $\phi(k)$ is also an important issue. When all $K_2$ vectors are weighted equally, that is, $\phi(0, k_2) = 1/K_2$, for $k_2 = 0, 1, \cdots, K_2 - 1$, then a causal rectangular window is defined. However, a causal exponential window is more useful because it emphasizes the recent rather than the distant estimates of the DVF [33]. Such a window is defined by

\[
\phi(k_1, k_2) = \begin{cases} \frac{1}{\mu} \exp \left( -\frac{k_1}{\mu} \right) & \text{if } k_1 = 0 \text{ and } k_2 \geq 0 \\ 0 & \text{if } k_1 \neq 0 \text{ and } k_2 < 0 \end{cases}
\tag{46}
\]

where the constant $\mu$ is equal to a number of samples. It is easily shown that the following simple recursion

\[
H(s, r + k) = \exp \left( -\frac{1}{\mu} \right) H(s, r) + \frac{1}{\mu} e(r) d(r - s)^T
\tag{47}
\]

for $k = (0, 1)$ is equivalent to (45) with $K_1 = 1$ and $K_2 \to \infty$. Examination of (45) reveals that each element of the matrix $H(s, r)$ is the 1-D convolution of two functions (gradient error data and weight coefficients), which is equivalent to the multiplication of their transforms in the discrete frequency domain. The Fourier transform of the exponential window is that of a single-pole lowpass filter with break frequency $1/\mu$ and its frequency response contains smaller side lobes than the rectangular window. In other words, as the 1-D exponential window is applied along the horizontal direction, it filters the error gradient data, with the goal of reducing the amount of noise present in the signal used for adaptation of the prediction coefficient. In tuning the system, varying $\mu$ is equivalent to varying the frequency response of the gradient estimation process.

Finally, it should be noted that although convergence of the LMS algorithm has not been demonstrated in the general case of decorrelated and nonstationary inputs [32], our experiments due to the lowpass nature of the DVF the adaptive algorithm converged for a value of $\beta$ within the bounds given by (44).

V. EXPERIMENTAL RESULTS

In this section, various implementation issues are discussed along with a number of experimental results that demonstrate the performance of the DVF estimation algorithms are presented. A recursive estimation scheme shown in Fig. 4 was used. According to the figure, frame $f_i(m, n)$ is scanned line by line and at each pixel the estimation of DVF is controlled based on the value of the DFD. A suitable criterion for terminating the iteration is obtained by using $\Xi(d - d^i) = 0$, which represents a measure of the average DFD in the mask $\mathcal{M}$. Then, following the initialization stage, where $d^i$ is computed by (31), a detector decides whether the working pixel belongs to the same moving area with its neighbors based on $\Xi$. If $\Xi \leq T$ holds, where the value of $T$ is computed based on the average squared frame difference or is adjusted experimentally, then, we set $d = d^i$ and no estimation of $u$ is required. Otherwise, $u$ is estimated and the updated
The displacement solution is given by

\[ \tilde{d}_{i+1}^* = \tilde{d}_i^* + \tilde{u}. \]  

The final displacement estimate is chosen to be \( \tilde{d}_{i+1}^* \) if inequality \( \Xi_{i+1}^* \leq \Xi_i^* \) is true, otherwise the previous estimate \( \tilde{d}_i^* \) is chosen. Note that, inequality \( \Xi_{i+1}^* \leq \Xi_i^* \) is useful in detecting discontinuities in the DVF, such as the transition of the recursion from moving to stationary areas. Thus, this scheme with the adaptive DVF prediction can provide a very efficient estimation of the DVF.

In order to evaluate the intensity value of the image \( f_{k-1} \) and the second-order spatial derivatives of \( f_{k-1} \) at an arbitrary location \( r = [x \ y]^T \), the following bilinear interpolation was used. The spatial gradient vector at \( r \) was approximated using backward differences as follows:

\[ \nabla f_{k-1} = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} (1 - \theta_i)(f_{10} - f_{00}) + \theta_i(f_{11} - f_{10}) \\ (1 - \theta_j)(f_{01} - f_{00}) + \theta_j(f_{11} - f_{10}) \end{bmatrix} \]  

(49)

where \( f_{ij} = f_{k-1}(i \cdot I, j \cdot Y) \). According to our experimental results, (49) provides improved estimates of the spatial gradient vectors and the DVF, compared to the case of setting \( \theta_i = \theta_j = 1/2 \) for all \( (x, y) \), as was done in [5], [7], and [10].

The proposed algorithms were tested on a standard video-conference monochromatic (8-b) image sequence “Trevor White” of size 150 \( \times \) 256 \( \times \) 256. Fig. 5 shows the 36th frame from this sequence. Since the original DVF is usually not known, a suitable figure of merit for the estimation algorithms is the improvement in motion compensation at each frame, which is defined in decibels (dB) by

\[ I_{MC}(k) = 10 \cdot \log_{10} \left( \frac{\sum_r \Delta_i^2(r, d)}{\sum_r \Delta_i^2(r, d - \tilde{d})} \right) \]  

(50)

where the summation is carried out for all spatial locations \( r = (m, n) \) on the \( k \)th frame. Note that \( I_{MC}(k) \) for each frame represents the ratio (in decibels) of the mean-squared FD over the mean-squared DFD based on the estimated DVF and is defined only in the presence of motion. When the entire sequence of \( K = 150 \) frames is considered (\( k = 1, 2, \cdots, K \)), then the average improvement in motion compensation is defined in decibels by

\[ I_{MC} = 10 \cdot \log_{10} \left( \frac{\sum_{k=2}^{K} \sum_r \Delta_k^2(r, d)}{\sum_{k=2}^{K} \sum_r \Delta_k^2(r, d - \tilde{d})} \right). \]  

(51)

Fig. 6 shows the values of \( I_{MC}(k) \), for frames 22–40 of the “Trevor White” sequence, obtained by four algorithms. First, the nonadaptive estimation algorithm given by (17) with \( W = I, \nu = 1, F_d = I \), and \( a = 0 \) was used (\( I_{MC} = 5.471 \) dB). This algorithm uses the least available prior knowledge about the uncertainty in the observations (bound \( \varepsilon^2 \)) and in DVF (bound \( E^2 \)), and it is equivalent to the Wiener-based algorithm proposed in [8]. The second algorithm results from the general algorithm given by \( (17) \) with \( W = I, \nu = 1, F_d = \Omega^{-1} \) and the stationary AR model for the DVF given in (22) (\( I_{MC} = 5.837 \) dB). In other words, by considering an unweighted single-input approach, the algorithm uses the same prior information about the observations as the first algorithm while it uses the OSC concept in combination with the AR model for the DVF as prior information about the solution. The third curve (\( I_{MC} = 6.257 \) dB) was obtained by considering a weighted multiple-input approach for the observations and the same stationary AR model as the previous algorithm. The correlation coefficients of the AR model were computed from the estimated DVF, resulting in the values \( \rho_{11} = 0.9678, \rho_{12} = 0.9684, \rho_{21} = 0.9872, \) and \( \rho_{22} = 0.9910 \), whereas the estimated variances of the displacement components were \( \sigma^2 = 0.2883 \) and \( \sigma^2 = 0.2503 \). Finally, the fourth algorithm used the adaptive weighted LMS approach described in Section IV in modeling the estimated DVF, where the value of \( \beta \) in (39) was adjusted.
experimentally to 0.018 and the value of $\mu = 5$ was used in (47) ($I_{MC} = 6.682$ dB). In all cases, the scheme given in Fig. 4 was used with $T = 0.2$. With the comparison shown in Fig. 6 the effect of each of the factors that modify the (Wiener-based) pel-recursive algorithm is demonstrated. These factors are 1) the weighting of the observations according to the estimated variance of the linearization error at each pixel, 2) the incorporation of prior knowledge about the solution in the form of an AR model of the DVF, 3) the weighting of the prior knowledge with the use of the OSC, and 4) the updating of the coefficients of the AR model. The effect of factors 1) and 3) has also been demonstrated in [12], [13], where it was shown that due to them the performance of the pel-recursive algorithm improves considerably. According to Fig. 6, the adaptive multiple-input algorithms using the AR model of the DVF and matrix $F_d$ outperform the nonadaptive (or Wiener) algorithm. In addition, the nonstationary adaptive modeling of the DVF also showed an improved performance compared to the stationary modeling approach. It is mentioned here that although the recursive scheme of Fig. 4 allows for iterative evaluation of $\hat{u}$ at each pixel, only one iteration was run for the results shown in Fig. 6 since no additional improvement was obtained by considering more iterations. This clearly constitutes an advantage of the proposed algorithm.

In order to observe the value of the MC prediction error that is achieved by the algorithms, the estimated DFD is mapped for display purposes according to

$$\Delta(r, d - \hat{d}) = \max \left\{ 0, \min \left( 128 + \eta \cdot \Delta_r (r, d - \hat{d}), 255 \right) \right\}$$

(52)

where $\eta$ is an appropriate scale factor ($\eta = 6$ was used in the following figures). Note that $\Delta(r, d)$ represents the (mapped) frame difference, which is shown in Fig. 7 for frames 35 and 36. Figs. 8 and 9 show, respectively, the values of $\Delta(r, d - \hat{d})$ based on the estimates of the DVF using the first and fourth algorithms as they were presented above and in Fig. 6. By comparing the two figures, it is clear that the reduction of DFD in Fig. 9 is considerably larger than that of Fig. 8, especially around the edges of the moving object. This property of the proposed algorithm is very important in the motion-compensated predictive coding of image sequences.

Although $I_{MC}$ is a useful figure of merit regarding the quality of the estimated DVF, it does not offer enough information about the smoothness of the estimated DVF. Due to the ill-posedness of the DVF estimation problem and the noise present, a maximally smoothed estimated DVF does not necessarily minimize the DFD in a given image sequence. However, in many applications, such as MC interpolation or MC prediction, the smoothness of the DVF is an important feature. Figs. 10 and 11 show a section of size 128 $\times$ 128 extracted from the center of the estimated DVF using the nonadaptive algorithm ($\nu = 1$, $W = I$, $F_d = I$, $a = 0$) and the adaptive algorithm ($\nu = 4$, $W$, $F_d = \Omega^{-1}$) with the nonstationary AR model, respec-

Fig. 6. $I_{MC}(k)$ for the "Trevor White" sequence.

Fig. 7. Frame difference $\Delta(r, d)$ for frames 35 and 36.

Fig. 8. $\tilde{\Delta}(r, d - \hat{d})$ obtained with $W = I$, $\nu = 1$, $F_d = I$, and $a = 0$. 
stationary and a nonstationary AR model, which is spatially adaptive according to the proposed weight LMS algorithm, have been presented. Such models have also been used in developing recursive filters for removing the noise from the estimated DVF [34]. In addition, a spatially adaptive DVF prediction approach is proposed for obtaining an initial estimate of the solution, as well as an "oriented smoothness" constraint for regularizing the ill-posed inverse problem to be solved. Due to its adaptivity, the performance of the algorithm is improved without increasing the size of the spatial mask. The latter is particularly important especially around occlusion areas of the image sequence. Although the computational load of the estimation is increased, only the inversion of $2 \times 2$ matrices is required regardless of the mask size. The algorithms presented here were compared experimentally with existing algorithms in estimating the DVF of standard video-conference image sequences. The performance of the proposed algorithms are shown to be superior to that of other simpler algorithms with respect to the reduction of the prediction error, robustness, as well as accuracy and smoothness of the estimated DVF.

**APPENDIX:**

**VARIANCE OF THE LINEARIZATION ERROR**

Let us consider the second-order terms of the Taylor's expansion in (3). Then, the linearization error at point $r_t$ in mask $\mathcal{A}_k$ is equal to

$$v_k(r_t, u(r_t)) = \frac{1}{2} u'(r_t)^\top B(r_t - \tilde{d}) u(r_t) + \epsilon_k(r_t, u(r_t)).$$

(A1)

where $u(r_t) = [u_1(r_t) u_2(r_t)]^\top$ is given by (1), $B(r_t - \tilde{d}) = \nabla \nabla f_{k-1}(r_t - \tilde{d})$ is the symmetric $2 \times 2$ matrix of second-order derivatives of $f_{k-1}$ at $r_t - \tilde{d}$, and $\epsilon_k(r_t, u(r_t))$ denotes the sum of the terms of order higher than two. Let us assume that the linearization term is a random

VI. CONCLUSIONS

In this paper a constrained gradient-based recursive DVF estimation algorithm was derived by following a set theoretic formulation. The estimated DVF is obtained recursively by minimizing at each pixel the linearized DFD using $\nu$ overlapping submasks that belong to a causal mask. Prior knowledge about the linearization noise is provided by considering an appropriate weight matrix for the observations, whereas prior knowledge about the solution is provided by an AR model of the DVF. Both a
variable with zero mean. Then, its variance is equal to
\[ c_{t}(l) = E\{v_{t}(r, u) v_{t}(r, u)^{T}\}. \]  
\hspace{1cm} \tag{A2}

In deriving a simple expression for \( c_{t}(l) \), we assume that \( \epsilon_{t} \) is uncorrelated with each element of \( B \). Due to the recursive estimation of the DVF, estimates of \( d_{t}(r) \) in the mask are available, therefore, based on (1), we have \( \hat{u}(r) = \hat{d}(r) \). Note that, \( \hat{d}(r) = 0 \) since the only prior available estimate of \( d_{t}(r) \) is \( d_{t} \). Then, an estimate of \( \hat{c}_{t}(l) \) is given by
\[ \hat{c}_{t}(l) = \frac{1}{4} E\{ \hat{u}(r) \hat{u}(r)^{T} \hat{d}(r) \hat{d}(r)^{T} \hat{d}(r) \hat{d}(r)^{T} \} + \hat{c}_{t}(l) \]  
\hspace{1cm} \tag{A3}

where \( \hat{c}_{t}(l) \) is an estimate of the variance of \( \epsilon_{t}(r, u) \).

After expanding the term \( H(l) \) we get
\[ H(l) = \sum_{(i, j, k, l)} E\{ \beta_{i}(r_{i}, d_{i}) \beta_{j}(r_{j}, d_{j}) \beta_{k}(r_{k}, d_{k}) \beta_{l}(r_{l}, d_{l}) \} \]  
\hspace{1cm} \tag{A4}

where \( \beta_{i} \) represents the \( i \)th element of the \( B \) matrix, and there are 16 quadruplets \( (i, j, k, l) \) since each element of \( (i, j, k, l) \) takes values from \( (1, 2) \). However, the computational load for evaluating (A4) for every pixel in each frame is considerable. Therefore, the following simplification is proposed:
\[ h_{ij,k}(l) = E\{ \beta_{i}(r_{i}, d_{i}) \beta_{j}(r_{j}, d_{j}) \beta_{k}(r_{k}, d_{k}) \beta_{l}(r_{l}, d_{l}) \} \]  
\hspace{1cm} \tag{A5}

where \( c_{ijk}(l) \) is the cross-variance of the elements of matrix \( B(r_{i}, d_{i}) \). By letting
\[ c_{ijk}(l) = \begin{cases} \sigma_{ijk}^{2} & \text{if } (i, j) = (k, l) \\ 0 & \text{otherwise} \end{cases} \]  
\hspace{1cm} \tag{A6}

and \( \hat{c}_{t}(l) = \sigma_{t}^{2}E, \forall l \in \mathcal{A} \), (A3) takes the form
\[ \hat{c}_{t}(l) = \frac{1}{4} \sum_{(i, j)} \sigma_{ij}^{2} \hat{u}_{i}(r_{i})^{2} \hat{u}_{j}(r_{j})^{2} + \sigma_{t}^{2} \]  
\hspace{1cm} \tag{A7}

Note that (A7) does not require the computation of the variance of the second-order derivatives at each point but only an estimate of their variance for the entire image. Therefore, the required computational load is reduced significantly. The second-order derivatives are estimated using the Beaudet operators [14]. Finally, \( \sigma_{t}^{2} \) is approximated by considering the variance of all third-order spatial derivatives of the image after it is lowpass filtered to avoid noise amplification. The experimental values of \( \sigma_{t}^{2} \) are in the range \([0.5, 1.5]\).

REFERENCES


Serafim N. Efstratiadis (M’91) was born in Thessaloniki, Greece, on January 10, 1964. He received the Diploma degree from the Aristotle University, Thessaloniki, Greece, in 1986, and the M.S. and Ph.D. degrees from Northwestern University, IL, in 1988 and 1991, respectively, all in electrical engineering.

From 1987 to 1991 he was a teaching assistant and a research assistant at the Department of Electrical Engineering and Computer Science of Northwestern University. Since November 1991 he has been First Assistant at the Signal Processing Laboratory of the Swiss Federal Institute of Technology, Lausanne, Switzerland. He is also the Head of the Digital TV group with 10 Ph.D. students, which is involved in the design and development of advanced video coding systems. He has published several conference papers and journal articles on signal/image processing and his current interests are multidimensional signal processing, motion-compensated image sequence modeling and image/video coding and transmission. Dr. Efstratiadis is a member of the Technical Chamber of Commerce of Greece.

Aggelos K. Katsaggelos (S’80–M’85) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, GA, in 1981 and 1985, respectively, both in electrical engineering.

From 1980 to 1985 he was a Research Assistant at the Digital Signal Processing Laboratory of the Electrical Engineering School at Georgia Tech. He is currently an Associate Professor in the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL. During the 1986–1987 academic year he was an Assistant Professor at Polytechnic University, Department of Electrical Engineering and Computer Science, Brooklyn, NY. His current research interests include signal and image processing, processing of moving images, and computational vision. Dr. Katsaggelos is an Ameritech Fellow and a member of the Associate Staff, Department of Medicine, at Evanston Hospital. He is also a member of the Society of Photo-Optical Instrumentation Engineers (SPIE), the Steering Committee of the IEEE TRANSACTIONS ON MEDICAL IMAGING and the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE–CAS Technical Committee on Visual Signal Processing and Communications, the IEEE–SP Technical Committee on Image and Multidimensional Digital Signal Processing, the Technical Chamber of Commerce of Greece, and Sigma Xi. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and also the editor of Digital Image Restoration (New York: Springer-Verlag, 1991).