A Recursive Nonstationary MAP Displacement Vector Field Estimation Algorithm

James C. Brailean, Member, IEEE, and Aggelos K. Katsaggelos, Senior Member, IEEE

Abstract—In this paper, a recursive model-based algorithm for obtaining the maximum a posteriori (MAP) estimate of the displacement vector field (DVF) from successive image frames of an image sequence is presented. To model the DVF, we develop a nonstationary vector field model called the vector coupled Gauss–Markov (VCGM) model. The VCGM model consists of two levels: an upper level, which is made up of several submodels with various characteristics, and a lower level or line process, which governs the transitions between the submodels. A detailed line process is proposed. The VCGM model is well suited for estimating the DVF since the resulting estimates preserve the boundaries between the differently moving areas in an image sequence. A Kalman type estimator results, followed by a decision criterion for choosing the appropriate line process. Several experiments demonstrate the superior performance of the proposed algorithm with respect to prediction error, interpolation error, and robustness to noise.

I. INTRODUCTION

ACURATE displacement field estimation is crucial in many applications of image sequence processing. In particular, image sequence coding, interpolation, object tracking, and spatio-temporal motion-compensated filtering are all applications of image sequence processing that utilize the interframe redundancy characterized by the displacement field (see, for example, [10], [15], [24], [30], [31], [37], and [40]). In this paper, the displacement vector field (DVF) represents the 3-D motion of objects projected onto the image plane. Estimation of the DVF from successive image frames is a formidable problem. Two of the main sources of the difficulties in obtaining accurate estimates are the ill-posed nature of the problem and the nonstationarity of the displacement field. Attempting to estimate the two unknown displacement vector components from a single intensity variation leads to the ill-posedness [3]. The nonstationarity of the DVF results from different moving objects within the scene that cause discontinuities to occur at the object boundaries, combined with regions where the motion is undefined due to covered or uncovered parts of the moving scene. These difficulties are only further aggravated when noise is present in the image sequence. In this paper, we propose a robust recursive estimator that overcomes both of these difficulties and provides significant improvements in the DVF estimates from noisy or noise-free image sequences over current displacement field estimators.

Recently, several optical flow based techniques for estimating the DVF have been developed that attempt to preserve the motion boundaries present in the DVF. For instance, Efstratiadis and Katsaggelos developed pixel recursive (PR) algorithms that adapt to the discontinuities of the DVF by estimating the variance of the linearization error term at each pixel [18]–[20] and by considering more than one consecutive frame [16]. They have also incorporated stationary as well as nonstationary autoregressive (AR) models of the DVF into the prediction component of the motion estimator [17], [18], [20]. In the latter case, the prediction coefficients are switched depending on the locations of boundaries that separate the moving and nonmoving areas between adjacent frames. Markov random field (MRF) models have recently gained in popularity [1], [5], [14], [22], [26], [34], [35], [42] as an alternative approach for preserving the motion boundaries when estimating the DVF. Specifically, the observed displacement field is modeled using a Markov random vector field (MRVF), while the discontinuities or structure of the DVF is modeled using a binary MRF. Utilizing the equivalence between the Gibbs distribution and a Markov random field, Konrad and Dubois [34] solved for both the global MAP and minimum expected cost estimates of the DVF using stochastic relaxation techniques. A drawback of stochastic relaxation techniques, such as simulated annealing, is their large computational cost. Similar approaches using MRF's were developed in [22], [26] and recently in [42]. In all cases less computationally complex, yet suboptimal, deterministic algorithms were developed. However, these deterministic algorithms still require several iterations prior to convergence.

In this paper, we also take the approach of modeling both the DVF and its discontinuities using Markov random fields. However, unlike the above approaches which use noneausal models for the DVF, we derive a recursive MAP estimator for the DVF. Furthermore, we develop a dynamic vector coupled Gauss–Markov (VCGM) model for the DVF, which represents an extension of the AR models presented in [6] and [16]–[18]. Similar scalar versions of this model have been successfully employed in the restoration of degraded still images [11], [27], [28]. It is a two-layered random field consisting of upper and lower layers. The upper layer is the observed displacement field, which is composed of several dynamic models representing a variety of local characteristics. The lower or hidden layer (line process) is a Markov random field with a finite range whose purpose is to govern the transitions
from one stationary region to another. In developing the stochastic characterization of the line process, we incorporate many of the a priori assumptions presented in [22], [26], and [34]. However, we further expand these characterizations to accommodate the recursive nature of the proposed estimator, while also increasing the sophistication of the line process. Through the use of this model, we develop a computationally efficient recursive MAP estimator that has the ability to adapt to the abrupt changes in the DVF, unlike the approaches mentioned above, which use iterative techniques. Since the estimation of the DVF from the intensity field is a nonlinear problem, a linearization step in the development of the MAP estimator is required. We develop the coupled linearized MAP (CLMAP), which approximates the optimal MAP estimator. The CLMAP is developed using a reduced order state vector, which results in the less computationally complex reduced order model CLMAP (RCLMAP). The proposed RCLMAP algorithm provides very accurate estimates of the DVF, as well as very accurate segmentation of the DVF, as is shown experimentally.

The remainder of the paper is organized as follows. In Section II, the displacement field estimation problem is formulated. The development of the vector coupled Gauss–Markov model and the coupled linearized MAP estimator are presented in Sections III and IV, respectively. Various implementation issues are discussed in Section V. Experimental results are presented in Section VI that highlight the effectiveness of the models and algorithms developed in Sections III and IV. Finally, conclusions and directions of future work are discussed in Section VII.

II. PROBLEM FORMULATION

Let \( f_k(r) \) denote the image intensity of the \( k \)th frame at the spatial location \( r \), where \( r = [m, n]^T \), \( 1 \leq m \leq M, 1 \leq n \leq N \), and \( T \) denotes the transpose of a vector or matrix. The displacement field \( \{d(x, r, k), d_y(r, k)\}^T \) is a vector field that maps points in the current frame \( f_k \) to their corresponding locations in the previous frame \( f_{k-1} \). If there is no possibility of confusion as to which pair of frames the displacement field belongs, then it is simply denoted as \( d(r) \). We impose the commonly used assumption that the image intensity is constant along the motion trajectory [40]. That is, knowledge of the true displacement vector allows for the perfect prediction of the next frame from the current one. We also assume that the image sequence has been corrupted by additive noise, that is, the \( k \)th frame \( f_k \) is given by \( f_k(r) = f'_k(r) + n_k(r) \), where \( n_k(r) \) represents zero-mean Gaussian noise with variance \( \sigma_n^2 \). Utilizing both the perfect registration assumption and the above expression for \( f_k(r) \) results in

\[
f_k(r) = f_{k-1}(r - d(r)) + N_k(r)
\]

where \( N_k(r) = n_k(r) - n_{k-1}(r - d(r)) \), \( \mathbb{E}[N_k(r)N_k(\hat{r})] = 2\sigma_n^2\delta(r - \hat{r}) \). \( \mathbb{E}[\cdot] \) denotes the expectation operator and \( \delta(\cdot) \) the delta function. Thus, the displacement field estimation problem is that of finding an estimate \( \hat{d}(r) = [d_x(r), d_y(r)]^T \) of the true displacement vector \( d(r) \) based on successive noisy frames.

It is apparent from (1) that the relationship between the DVF and the intensity field is a nonlinear one. For the development of the recursive estimator to be presented in Section IV, a linear relationship between the observed intensity values and the DVF will be required. Using a Taylor series expansion of \( f_{k-1}(r - d(r)) \) about the location \( r - d(r) \) where \( d^2(r) \) denotes an initial estimate of \( d(r) \), we obtain such an expression for the displaced frame difference (DFD) \( \Delta(r, r - d^2(r)) \)

\[
\Delta(r, r - d^2(r)) = f_k(r) - f_{k-1}(r - d^2(r)) = \nabla^T f_{k-1}(r - d^2(r))(d(r) - d^2(r)) - e_L(r, d^2(r)) + N_k(r)
\]

where \( e_L(r, d^2(r)) \) is the error resulting from the truncation of the higher order terms and \( \nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]^T \) represents the spatial gradient operator. We can interpret the DFD as the error due to the nonlinear temporal prediction of the intensity field. It is this error that motion estimation algorithms attempt to minimize. The direct minimization of the mean square error results in the steepest descent class of PR algorithms [38]-[40]. Minimizing the linearized form of the DFD with respect to \( d(r) \) results in an updated estimate of the displacement vector. Therefore, an estimate of \( d(r) \) results from knowledge of the DFD (temporal gradient) and the spatial gradients. The truncation error was ignored in earlier algorithms, while it was treated as a sample of a random field in [4], [16], [18], and [19]. In the MAP estimator we propose, we use both the linearized and the original expression for the DFD and assume that the combined error term \( e(r, d^2(r)) = e_L(r, d^2(r)) - N_k(r) \) is a Gaussian random variable with variance \( \gamma \). However, with the incorporation of the VCGM model and MAP criteria, the resulting estimator has a significantly different form from the above mentioned algorithms.

III. VECTOR COUPLED GAUSS–MARKOV (VCGM) MODEL

A. The Dynamic Model

The modeling of the DVF potentially improves the robustness of the estimation process. For instance, the linearization error term \( e_L \) is directly affected by the accuracy of the initial estimate \( d^2 \). By incorporating a model that utilizes the spatial correlations of the DVF, an initial estimate of the displacement vector can be obtained, which minimizes a prediction error function [18]. However, as is the case with the modeling of the image intensity field, the DVF is not modeled properly by a linear shift invariant (LSI) model. The major problem with using an LSI model in the DVF estimation process is that the boundaries between differently moving or moving and stationary areas will become over-smoothed or blurred. Therefore, as already mentioned in the introduction, we propose modeling the DVF using a vector coupled Gauss–Markov (VCGM) model. Through the use of the lower layer or line process, we estimate the boundaries or structure of the DVF and thus avoid the errors associated with the over-smoothing of the boundaries.
The components of the line process are discrete random variables called line elements. More specifically, \( l(r, k) \) is defined as a set of line elements

\[
l(r, k) = \{l_{m,n,k}(i, j)\}
\]

where each line element \( l_{m,n,k}(i, j) \) defines the presence \((l_{m,n,k}(i, j) = 1)\) or absence \((l_{m,n,k}(i, j) = 0)\) of a boundary between displacement vectors \( d(m, n, k) \) and \( d(m-i, n-j, k) \). As was done for the DVF, if there is no possibility of confusion as to which pair of frames the line process belongs, then it is simply denoted as \( l(r) = \{l_{m,n}(i, j)\} \). The VCGM model results from the coupling of this line process with several vector Gauss–Markov (VGM) models. It is given by

\[
d(r) = \sum_{i,j \in \mathbb{Z}} a_{i,j}^{(r)} d(m-i, n-j) + w^{(r)}(r)
\]

where \( a_{i,j}^{(r)} \) is a 2 × 2 matrix containing the prediction coefficients, \( w^{(r)}(r) \) is a zero mean Gaussian noise component with covariance matrix \( \Phi^{w(r)}(r) \), and \( \mathbb{Z} \) denotes the nonsymmetric half plane (NSHP). Without loss of generality, in the following, we assume that the support of the NSHP model extends one pixel to left, right, and above. The line elements associated with the \( m, n \) pixel are shown in Fig. 1 (solid lines). The dependence of \( a_{i,j}^{(r)} \), \( w^{(r)}(r) \), and \( \Phi^{w(r)}(r) \) on \( l(r) \) can be explained as follows. With knowledge of \( l(r) \), a model is chosen that excludes the displacement vectors that are separated by a discontinuity. This is accomplished by setting to zero the prediction parameter \( (a_{i,j}^{(r)}) \) that corresponds to an activated line element \((l_{m,n}(i, j) = 1)\), which effects the uncertainty \( (w^{(r)}(r)) \) in the model’s accuracy. The region of support was chosen to be the nonsymmetric half plane to allow recursive compatibility. It is noted here that the VCGM model can be expanded to three dimensions, as was done for a stationary AR model of the DVF in [18]. However, a 3-D model is not used here since we take advantage of the temporal correlations that exist in the DVF, indirectly, through the use of an energy function, as explained in the next section.

B. The Mixed Joint Probability Density Function

Since the ultimate goal is to determine the MAP estimates of the displacement field and line process, knowledge of the \( a \) priori mixed joint probability density function (mjpdf) of \( d(r) \) and \( l(r) \) is required. Noting that \( d(r) \) and \( l(r) \) are jointly Markovian, we know from the Gibbs equivalence [21], [34] that they are jointly Gibbsian. Thus, the mjpdf of \( d(r) \) and \( l(r) \) is given by

\[
\rho(d, l) = \frac{1}{Z} \exp \left( -\frac{U_d(d \mid l) + U_l(l)}{\beta} \right)
\]

(5)

where \( Z \) is a normalizing constant, independent of \( d \) and \( l \). In the derivation of the MAP estimator to be presented in Section IV, the conditional mjpdf \( \rho(d, l \mid f_{k-1}) \) will be required, where \( f_{k-1} \) represents the entire previous frame. It is equal to

\[
\rho(d, l \mid f_{k-1}) = \frac{1}{Z} \exp \left( -\frac{U_d(d \mid l, f_{k-1}) + U_l(l \mid f_{k-1})}{\beta} \right)
\]

(6)

where the conditioning on \( f_{k-1} \) in \( U_d(d \mid l, f_{k-1}) \) is omitted since the estimation of \( d \) is independent of any particular frame and therefore \( f_{k-1} \). In other words, knowledge of \( f_{k-1} \) (or any other frame) alone does not determine \( d \). The conditional energy function \( U_d(d \mid l) \) is defined as

\[
U_d(d \mid l) = \sum_{i,j \in \mathbb{Z}} V_d(d \mid l)
\]

(7)

where the “potential functions” \( V_d(d \mid l) \) depend only on those samples \( d(r) \), which are contained within the local neighborhood \( \mathbb{Z} \). It is these potential functions that are crucial to the characterization of the DVF.

To determine the potential functions, we take advantage of the underlying properties of the models used to describe the DVF. That is, since the line process \( l(r) \) is given, we know that the resulting VGM model is characterized by a Gaussian pdf. Therefore, the potential functions are chosen to reflect this property, that is

\[
V_d(d \mid l) = \left\{ \begin{array}{ll}
\frac{1}{2} \sum_{i,j \in \mathbb{Z}} a_{i,j}^{(r)} d(m-i, n-j) \Phi^{-1}(a_{i,j}^{(r)}) d(m-i, n-j) & \text{if } (m, n) \in \mathbb{Z} \\
\frac{1}{2} d^*(m, n) \Phi^{-1}(a_{i,j}^{(r)}) & \text{if } p = q = 0.
\end{array} \right.
\]

(8)

Noting that the prediction coefficients are dependent on the line process, the role of the energy function \( U_d(d \mid l) \) is explained as follows. If the displacement field begins to deviate from the \( a \) priori assumptions provided by the VGM model, then the energy function \( U_d(d \mid l) \) is increased by the cost \( V_d(d \mid l) \). If, however, the line element is “activated” \((l_{m,n}(p, q) = 1)\), then the corresponding prediction coefficient \( a_{i,j}^{(r)} \) is set to zero. Thus, there is no cost associated with the pair of pixels \((m, n) \) and \((m-p, n-q) \). Therefore, the introduction of abrupt changes in the DVF is not penalized.

To prevent the line field \( l(r) \) from being on everywhere, the energy function for the line process \( U_l(l \mid f_{k-1}) \) is increased for each “activated” line element. However, we also use this term to incorporate additional knowledge that is available concerning the line process. Therefore, with knowledge of the states of the past spatial as well as temporal line elements,
we encourage continuity through the neighborhood, while penalizing crossings, isolated discontinuities, parallel lines, and inconsistencies in the line process. The existence of secondary correlation paths around an “activated” line element result in an estimated line field that is inconsistent. These paths are generally a result of the corner elements \( l_{m,n}(1,1) \) and \( l_{m-1,n}(1,-1) \).

To discourage the above traits, while at the same time encouraging continuity in the line field, we propose the following energy function:

\[
U_l(\mathbf{r} \mid f_{k-1}) = \alpha_t \sum_{i,j \in \mathbb{R}^2} l_{m,n}(i,j) + \alpha_c \sum_{i,j \in \mathbb{R}^2} l_{m,n}(i,j) \\
+ \alpha_c \left\{ \sum_{j=1}^{l_{m,n}(1,1)} l_{m,n}(1,j)(1 - l_{\text{edge}}(m,n)) \\
+ l_{m,n}(0,1)l_{m,n-1}(0,1) \right\} \\
+ \alpha_g \sum_{j=1}^{l_{m,n}(1,1)} l_{m,n}(1,j)(1 - l_{\text{edge}}(m,n)) \\
- l_{\text{edge}}(m - 1, n - 1) \} \\
+ \alpha_d U_d(\mathbf{r} \mid f_{k-1}) + \alpha_c U_c(\mathbf{r} \mid f_{k-1})
\] (9)

where \( l_{\text{edge}}(m,n) \) is a binary variable indicating the presence \( l_{\text{edge}}(m,n) = 1 \) or absence \( l_{\text{edge}}(m,n) = 0 \) of an intensity edge in \( f_k \) and the line element \( l_{m,n}(i,j) \) is obtained by motion compensating the previous line field using \( \mathbf{d}^* (\mathbf{r}, \mathbf{k}) \). That is

\[
l_{m,n}(i,j) = l_{m,n}(i,j) - \mathbf{d}^* (\mathbf{r}, \mathbf{k}).
\] (10)

The various weights are determined experimentally as is discussed in Section V-C. The first term of (9) is the cost associated with the introduction of a line element. More specifically, through this term we take advantage of the knowledge that the displacement field is generally smooth. Therefore, a cost is assigned for each activated line element. The second term penalizes each introduction of a line element that does not correspond to its motion compensated line element in the previous DVF. Therefore, we encourage temporal continuity between displacement fields by penalizing the sudden appearance or disappearance of discontinuities in the DVF. The third term prevents the formation of parallel lines. With respect to the fourth term, we know that in general, a motion boundary should correspond to an intensity edge (although an intensity edge does not necessarily correspond to a motion boundary). Therefore, we assume that an “activated” line element should coincide with an intensity edge. This assumption is characterized by the fourth term, where the cost of activating a line element is dependent on the spatial edge map of \( f_k \), which is found using any one of the many edge detection algorithms available. Finally, the last term encourages the formation of continuous contours of discontinuities and penalizes the formation of isolated line elements and inconsistencies in the line field. The terms \( U_l(\mathbf{r} \mid f_{k-1}) \) and \( U_d(\mathbf{r} \mid f_{k-1}) \) represent the energy functions for the continuity and consistency traits, respectively. That is, through \( U_d(\mathbf{r} \mid f_{k-1}) \), we encourage the formation of continuous line element contours, while penalizing crossings and isolated points. As was mentioned above, we enforce these costs through the neighborhood \( \mathbb{R}^2 \), taking into account the state of the surrounding line elements. Specifically, we look at the four line elements \( l_{m-1,n}(1,-1), l_{m,n-1}(1,0), l_{m-1,n}(0,1), \) and \( l_{m,n-1}(0,1) \), that surround the four elements within the neighborhood, which are known due to the recursive solution to be developed (dotted lines in Fig. 1). If any of these surrounding four line elements are “activated,” then continuity through the neighborhood is encouraged. The actual expression for \( U_d(\mathbf{r} \mid f_{k-1}) \) is dependent upon the values of the corner terms \( l_{m-1,n}(1,1) \) and \( l_{m,n}(1,1) \) and is given by

\[
U_d(\mathbf{r} \mid f_{k-1}) = l_{m,n-1}(1,-1)l_{m,n}(1,1) \\
- \{[l_{m,n}(0,1) - l_{m,n}(1,1)]^2 \\
- 2l_{m,n}(0,1)l_{m,n}(1,1)l_{m,n}(0,1)l_{m,n}(1,1) \\
+ l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,1)[l_{m,n}(1,1) - l_{m,n}(1,1)]^2 \\
+ l_{m,n-1}(1,-1)l_{m,n}(1,1) \} \\
2[l_{m,n}(0,1)l_{m,n-1}(0,1)l_{m,n}(1,0)l_{m,n}(1,1) \\
+ l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,1)[l_{m,n}(1,1) - l_{m,n}(1,1)]^2 \\
+ l_{m,n-1}(1,-1)l_{m,n}(1,1) \} \\
- [l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,0)l_{m,n}(1,1) \\
+ l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,1)[l_{m,n}(1,1) - l_{m,n}(1,1)]^2 \\
+ l_{m,n-1}(1,-1)l_{m,n}(1,1) \} \\
- [l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,0)l_{m,n}(1,1) \\
+ l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,1)[l_{m,n}(1,1) - l_{m,n}(1,1)]^2 \\
+ l_{m,n-1}(1,-1)l_{m,n}(1,1) \} \\
- [l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,0)l_{m,n}(1,1) \\
+ l_{m,n}(1,l)l_{m,n-1}(0,1)l_{m,n}(1,1)[l_{m,n}(1,1) - l_{m,n}(1,1)]^2 \\
+ l_{m,n-1}(1,-1)l_{m,n}(1,1) \} \right) \\
= l_{m,n-1}(1,-1)l_{m,n}(1,1)
\] (11)

where \( l_{m,n}(1,1) \) is the complement of \( l_{m,n}(1,1) \), and \( \infty \) is the symbol of infinity. From this equation, we see that \( U_d(\mathbf{r} \mid f_{k-1}) \) is broken up into four major terms (terms inside the brackets), each being weighted by a different possible combination of \( l_{m,n-1}(1,1) \) and \( l_{m,n}(1,1) \). As an example, let both \( l_{m,n-1}(1,1) \) and \( l_{m,n}(1,1) \) be “activated” (see Fig. 1), thus forcing the three other possible combinations to zero. If a continuous horizontal or vertical line passes through these elements and the neighborhood, then no cost is paid. However, if both the horizontal and vertical lines are present, or if a line terminates in the neighborhood, then there is a cost associated with each of these situations. The cost of a particular line configuration is dependent upon the number of violations detected.
when analyzed with the surrounding line elements. Similar examples exist for other combinations of \(l_{m,n-1}(1,-1)\) and \(l_{m,n}(1,1)\). It should be noted, however, that the last term in \(U_i(r | f_{k-1})\) does not have an incremental cost structure. This is due to our knowledge that any configuration occurring with both \(l_{m,n-1}(1,-1)\) and \(l_{m,n}(1,1)\) off requires objects of one pixel width. Since the probability of this occurring is small for most types of sequences, we unilaterally discourage any configuration (except for the case when all elements are off) with this characteristic from forming. However, if the a priori knowledge about the sequences under consideration changes, this term could be modified to reflect this knowledge.

As the name implies, the consistency energy term \(U_i(r | f_{k-1})\) assigns costs to a particular line configuration that violates the consistency of the line field. As was mentioned above, this occurs when a secondary correlation path exists around a point of discontinuity. An example of this occurs when \(l_{m-1,n}(0,1), l_{m,n}(1,1),\) and \(l_{m,n}(0,1)\) are on (see Fig. 1), while all the other line elements in the \(R\) neighborhood are off, giving the appearance of a vertical line through the neighborhood. However, since \(l_{m,n-1}(1,1)\) and \(l_{m,n}(0,1)\) are off, there exists a path of correlation between pixels \((m,n-1)\) and \((m,n)\) and another between \((m-1,n)\) and \((m,n)\), thus establishing a secondary correlation between pixels \((m,n-1)\) and \((m,n)\). Therefore, the discontinuity indicated by \(l_{m,n}(0,1)\) is circumvented by this secondary path, making the field inconsistent. To discourage these types of inconsistencies from forming in the line field, \(U_i(r | f_{k-1})\) is defined as follows:

\[
U_i(r | f_{k-1}) = l_{m,n-1}(1,-1)l_{m,n}(1,1) \\
\cdot \left\{ [l_{m,n}(0,1)l_{m-1,n}(0,1) - l_{m-1,n}(1,0)l_{m,n-1}(1,0)]^2 \right. \\
+ \left. [l_{m,n}(0,1) + l_{m-1,n}(0,1) + l_{m,n}(1,0) + l_{m-1,n}(0,1)]^2 \right\}
\]

\[
\cdot \left[ l_{m,n}(1,0) - l_{m,n-1}(1,0) - l_{m-1,n+1}(0,1) \right]^2 \\
- l_{m,n}(1,0)l_{m,n}(n-1,1)l_{m,n-1}(1,0) \\
+ l_{m,n}(n-1,1)l_{m,n}(1,0) \\
- \left[ l_{m,n}(0,1)l_{m-1,n}(0,1)l_{m,n}(1,0)l_{m,n-1}(1,0) \right] \\
+ \left[ l_{m,n}(0,1) - l_{m,n-1}(0,1) \right]^2 \\
- l_{m,n}(1,0) - l_{m,n}(1,1) - l_{m,n}(1,0)l_{m,n-1}(1,0) \\
+ l_{m,n}(1,0)l_{m,n}(1,1) \\
- l_{m,n}(0,1)l_{m,n}(1,1)l_{m,n-1}(1,0) \\
+ l_{m,n}(0,1)l_{m,n}(1,0)l_{m,n-1}(1,0) \\
\cdot \left\{ 2[l_{m,n}(0,1)l_{m-1,n}(0,1)l_{m,n}(1,0)l_{m,n-1}(1,0)]^2 \right. \\
+ \left. [l_{m,n}(0,1) - l_{m,n-1}(0,1)]^2 \right. \\
- l_{m,n}(1,0) - l_{m,n}(0,1)l_{m,n-1}(1,0) \\
+ \left. l_{m,n}(1,0) - l_{m,n}(1,0)l_{m,n-1}(1,0) \right]^2
\]  
(12)

\(U_i(r | f_{k-1})\) is also broken up into three major terms, each being weighted by different combinations of \(l_{m,n-1}(1,-1)\) and \(l_{m,n}(1,1)\). However, we no longer carry the term corresponding to the case when both the corner elements \(l_{m,n-1}(1,-1)\) and \(l_{m,n}(1,1)\) are off, since these configurations are strongly discouraged in \(U_i(r | f_{k-1})\).

As previously mentioned, the proposed energy function \(U_i(r | f_{k-1})\) for the line process has conceptual similarities to those presented in [22], [26], [34], and [42]. However, the proposed \(U_i(r | f_{k-1})\) is distinctly different in that a) it is formulated using only the causal line element neighbors surrounding the pixel under consideration, b) we take into account the interactions between the diagonal neighbors of the DVF, which are not considered by the other simpler models, and c) we also include in our assumptions of the line process the idea of temporal continuity between line fields and consistency, which are not considered in the other formulations of \(U_i(r | f_{k-1})\).

IV. COUPLED LINEARIZED MAP (CLMAP) ESTIMATOR

In this section, we develop an estimator for the displacement field \(d(r)\) and line process \(l(r)\) by maximizing the joint a posteriori probability density function, with respect to both \(d(r)\) and \(l(r)\). We begin our development by defining a reduced order state displacement vector as

\[
S(r) = [d^T(m,n), d^T(m,n-1), d^T(m-1,n+2), \\
\quad \quad \quad \quad d^T(m-1,n), d^T(m,n)]^T.
\]

(13)

The components of \(S(m,n)\) are chosen such that \(S(m,n)\) contains the proper elements to satisfy the NSHP region of support for the model. The above definition allows for the DVF to be expressed as a first order dynamic system, according to

\[
S(m,n) = A_i^{(r)} S(m,n-1) + B_i u(m,n) + Z_i V_i^{(r)}(m,n).
\]

(14)

The input variable \(u(m,n)\) is defined to be the most recent estimate of the displacement vector at the spatial location \((m-1,n+2)\), that is

\[
u(m,n) = \hat{d}(m-1,n+2) + \varepsilon
\]

(15)

where \(\varepsilon\) represents the uncertainty in the estimate \(\hat{d}(m-1,n+2)\). Thus, the noise term \(V_i(m,n)\) includes the original spatially varying plant noise \(w_i^{(r)}(m,n)\) together with \(\varepsilon\)

\[
V_i^{(r)}(m,n) = [w_i^{(r)}(m,n) \varepsilon]^T
\]

(16)

and

\[
Z = \begin{bmatrix}
I & 0 & 0 & 0 & 0 \\
0 & I & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

where \(I\) is a \(2 \times 2\) identity matrix and \(0\) is a \(2 \times 2\) zero matrix.

The remaining model matrices are defined as follows:

\[
A_i^{(r)} = \begin{bmatrix}
a_{i0}^{(r)} & 0 & a_{i1}^{(r)} & a_{i0}^{(r)} & a_{i1}^{(r)} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]
and

$$B = [0 \ 0 \ 1 \ 0 \ 0]^T.$$  

We assume that both components in the noise term $V(m,n)$ are independent, which results in the following covariance matrix

$$P_V = \begin{bmatrix} \Phi_{w(t)} & 0 \\ 0 & P_r \end{bmatrix}.$$  

(17)

Using the definition for $S(r)$ in (13), the mixed joint a posteriori probability density function for $S(m,n)$ and the corresponding $l(m,n)$ is now expressed, using Bayes rule, as

$$\rho(S(r), l(r) \mid f_k(r), f_{k-1}) = \frac{\rho(f_k(r) \mid f_{k-1}, S(r), l(r))\rho(S(r), l(r) \mid f_{k-1})}{\rho(f_k(r) \mid f_{k-1})}.$$  

(18)

The objective is to find estimates $\hat{S}(r)$ and $\hat{l}(r)$ such that $\rho(S(r), l(r) \mid f_k(r), f_{k-1})$ is maximized. Note that since $\rho(f_k(r) \mid f_{k-1})$ is not a function of $S(r)$ or $l(r)$, it is ignored during this maximization process. From (2) and the assumption that the combined error term $e(r, d')$ is Gaussian, the conditional density $\rho(f_k(r) \mid f_{k-1}, S(r), l(r))$ is Gaussian.

The a priori density $\rho(S(r), l(r) \mid f_{k-1})$ is given by (6), with $HS(r)$ replacing $d(r)$ where $H = [I, 0, \ldots, 0]$. Based on (6), we rewrite the a priori density as

$$\rho(S(r), l(r) \mid f_{k-1}) = \rho(S(r) \mid l(r)) P(l(r) \mid f_{k-1}).$$  

(19)

Recalling that VGM models are used to describe the upper layer of the DVF, we know from (4) that the conditional density $\rho(S(m,n) \mid l(m,n))$ is Gaussian. Therefore, if we assume that the line process is known $P(l(r) \mid f_{k-1}) = 1$, then the resulting density is clearly Gaussian. Furthermore, the recursive, MAP estimator has the form of an extended Kalman filter [36]

$$S^-(m,n) = A^{l(r)}S^+(m,n-1)$$  

(20)

$$E^-(m,n) = A^{l(r)}E^+(m,n-1)A^{l(r)T} + Q^{l(r)}$$  

(21)

$$\hat{S}^+(m,n) = \hat{S}^-(m,n) + K(m,n) \times [f_k(m,n) - f_{k-1}(r - HS^{-1}(r))]$$  

(22)

$$K(m,n) = E^-(m,n)G(r - HS^{-1}(r)) \times [G^T(r - HS^{-1}(r))E^-(m,n)] + \gamma^{-1}$$  

(23)

$$E^+(m,n) = [I - K(m,n)G(r - HS^{-1}(r))]E^-(m,n)$$  

(24)

where the superscripts $(-, +)$ indicate the estimate before and after updating with the innovation process, respectively. $E^+(m,n)$ represents the covariance of the error associated with the estimate $S(m,n)_{i=\{\ldots, +\}}$, $G(r - HS^{-1}(r)) = -H^T\nabla f_{k-1}(r - HS^{-1}(r))$, $\gamma$ is the variance of the combined error term $e(r, HS^{-1}(r))$ resulting from (2), and $Q^{l(r)} = \sum P_V^{l(r)T}$. The above recursive MAP estimator is a reduced order model extended Kalman filter, since it is based on a reduced order state vector. The linearized observation model (2) is used in the derivation of the gain $K(m,n)$ (23), while the nonlinear form of the displaced frame difference is used for the innovations term (22). The scalar reduced order model Kalman filter (without a line process and with a linear observation model) was presented by Angwin and Kaufman in [2], and ch. 7 of [31]. In their development, they showed that using a reduced order state vector with the optimal Kalman equations resulted in a large savings in the number of computations required, with minimal degradation in performance.

The assumption that we know $l(m,n)$ exactly is unrealistic. However, we do know that the number of possible configurations for the line process is limited to $L = 16$ for the NSHP of Fig. 1, and that the a priori probability of any one of these configurations is given by a Gibbs distribution with energy function defined in (9). Thus, the RCLMAP estimate of the DVF is determined by deciding which configuration of $l(m,n)$ and resulting estimate $\hat{S}^+(m,n)$ maximizes (18).

More specifically, the RCLMAP estimator is described as $L$ simultaneously operating extended Kalman filters, each using a different model that corresponds to a particular line process configuration. The line configuration providing the best estimate of $S(m,n)$ as determined by a MAP decision criterion is retained as the estimate $\hat{l}(m,n)$. A similar decision criterion was proposed in [25] and [44] for the restoration of degraded images. To determine the MAP decision rule for the line process, we note that the above estimator provides the estimate $\hat{S}_i$ such that $\rho(\hat{S}_i(m,n) \mid l_i(m,n))$ is maximized for each possible $l_i$. Therefore, we need to decide which $l_i$ maximizes (18). With the use of (19), this translates to determining $l_i$ which maximizes $\rho(f_k(r) \mid f_{k-1}, \hat{S}_i(r), l_i(r))P(l_i(r) \mid f_{k-1})$. Using the knowledge that $\rho(f_k(r) \mid f_{k-1}, \hat{S}_i(r), l_i(r))$ is Gaussian and $P(l_i(r) \mid f_{k-1})$ is Gibbsian, we express the product as

$$\rho(f_k(r) \mid f_{k-1}, \hat{S}_i(r), l_i(r))P(l_i(r) \mid f_{k-1}) = \kappa \exp\left\{\frac{-\Delta^2(r, r - HS^{-1}(r)) - U_i(l_i(r))}{2\sigma^2_i}\right\}$$  

(25)

where

$$\sigma^2_i = G^T(r - HS^{-1}(r))E^+(m,n)G(r - HS^{-1}(r)) + \gamma$$

$\kappa$ is a normalizing constant, and $\Delta_i$ is defined in (2). Therefore, from (25), the MAP decision rule is given by selecting $l_i(m,n)$ such that

$$\Delta^2_j(r, r - HS^{-1}(r)) \leq \min_{j=1 \ldots L} \Delta^2_j(r, r - HS^{-1}(r)) + c_j$$  

(26)

where

$$c_i = \frac{1}{2} \log \left\{\left|\sigma^2_i\right|\right\} + U_i(l_i(m,n)).$$

The above quantity used for comparison in the MAP decision rule can be described as the local total energy function, where the squared DFD is the energy of the estimation error and $c_i$ contains the energy function for the line process. Therefore, through the use of this decision rule, the estimates $d(r)$ and $l(r)$ that provide the minimum total energy are selected. Although we cannot claim that our estimate is the global MAP estimate, we do obtain a local MAP estimate, since at each location the minimum energy state is chosen. Therefore, the RCLMAP algorithm is dependent upon its initial conditions.
It should be noted that Kalman filters have been proposed in the past for estimating the DVF \cite{13}, \cite{43}. However, a stationary random field is assumed in the development of these approaches (no results were presented in \cite{13} on the effectiveness of this approach). A Kalman estimator was also formulated in \cite{12} for the restoration of the DVF. The above proposed RCLMAP estimator can be described as an extended Kalman filter that has the capability of changing its prediction coefficients abruptly in concert with the discontinuities of the DVF.

**A. Iterative RCLMAP**

There may be instances when the estimate of the DVF becomes less reliable. These instances generally occur when there is large interframe motion or at the junction of two boundaries. The final estimated \( d(\mathbf{r}) \) can be improved by implementing local iterations of the RCLMAP algorithm, where a local iteration is defined to be over a single pixel. To implement this iteration, we use the updated estimate \( \hat{S}_{c}^{+}(m, n) \) as the initial estimate for a second update estimate, and so on. We can define the iterative RCLMAP as a combination of the reduced order model and (20)–(24) with iterations performed on (22)–(24). More precisely, \( \hat{S}_{c}^{+}(m, n) \) becomes

\[
\hat{S}_{c}^{+}(m, n) = \hat{S}_{c}^{-1}(m, n) + \mathbf{K}_{c}(m, n) \times [f_{c}(m, n) - f_{k-1}(\mathbf{r} - \hat{H}_{c}^{+}(\mathbf{r}))]
\]

(27)

\[
\mathbf{K}_{c}(m, n) = \mathbf{E}_{c}^{-1}(m, n)G_{c}(\mathbf{r} - \hat{H}_{c}^{+}(\mathbf{r}))
\times [G^{T}(\mathbf{r} - \hat{H}_{c}^{+}(\mathbf{r}))\mathbf{E}^{-1}(m, n)]
\times \mathbf{G}(\mathbf{r} - \hat{H}_{c}^{+}(\mathbf{r})) + \gamma]^{-1}
\]

(28)

\[
\mathbf{E}_{c}^{+}(m, n) = [I - \mathbf{K}_{c}(m, n)\mathbf{G}(\mathbf{r} - \hat{H}_{c}^{+}(\mathbf{r}))]\mathbf{E}_{c}^{-1}(m, n)
\]

(29)

for \( i = 2, 3, \ldots, I \) and initial conditions \( \hat{S}_{c}^{+}(m, n) = \hat{S}^{+}(m, n) \), \( \mathbf{K}_{c}(m, n) = \mathbf{K}(m, n) \), and \( \mathbf{E}_{c}^{-1}(m, n) = \mathbf{E}^{+}(m, n) \). We obtain the final estimate for our reduced order state vector by setting \( \mathbf{d}^{+}(m, n) = \mathbf{d}^{+} \) where \( \mathbf{L} \) can be a preset number or when the iteration process is stopped due to some criteria. Note that \( i = 1 \) is just the RCLMAP estimator.

**V. IMPLEMENTATION ISSUES**

The model parameters for the VGM submodels were obtained from a prototype sequence using a standard least squares minimization technique over the estimated DVF. Since the relative spatial correlations of the DVF remain constant over large segments of an image sequence, the prediction coefficients for the VGM submodels need only be computed once. However, if the relative correlations of the DVF become temporally nonstationary (i.e., large scene changes occur), then the updating of the prediction coefficients is required.

The a priori probability distribution function for the line processes is described using a Gibbs distribution with energy functions described in (9). It is through these energy functions that we model characteristics of the line processes. Therefore, we are required to set several parameters \( (\alpha_{i}) \) that weight the importance of these characteristics. The same requirements are also found in the development of the other DVF and intensity field estimation techniques \cite{9}, \cite{22}, \cite{26}, \cite{34}, \cite{42}, which use the equivalence between MRF and a Gibbs distribution. In all cases, these cost parameters are found experimentally. However, the parameters were found to be consistent among similar sequences. Therefore, these parameters need only be computed once for classes of similar images with respect to motion and objects.

The \( \alpha_{i} \)'s were set according to the values determined in \cite{9}. In this work, it was found that the value of \( \alpha_{t} \), which is the weight associated with the cost of introducing a line element, should be set equal to the inverse of the variance of the DVF. Therefore, if the variance of the DVF is small, then the cost of activating a line element is large, while the inverse is true when the variance of the DVF is large. An initial estimate of the variance of the DVF is obtained from the normal equations used in estimating the prediction coefficients, as described above; it is then updated after each new DVF is estimated. The value of \( \alpha_{c} \), which is the weight associated with the continuity of the line process in the temporal direction, was set equal to the inverse of the previous squared DFD \[ f_{k-1}^{2}(x, y) - f_{k-2}^{2}(x, y) \times d^{2}(x, y, k-1) \] motion compensated by \( d^{2}(x, y) \). Therefore, \( \alpha_{c} \) is a spatially varying cost parameter. Note that if no previous DVF is available, then \( \alpha_{c} \) is set to zero. For the cost parameters \( \alpha_{p}, \alpha_{g}, \alpha_{d}, \) and \( \alpha_{t} \), no simple relationship with the estimated DVF's were detected from the experiments. These weights were chosen under the assumption that the corresponding characteristics of the line process are of equal importance. In other words, they were determined according to the range of values of the corresponding energy terms. For example, since \( 0 \leq U_{d}(r \mid f_{k-1}) \leq 3 \) and \( 0 \leq U_{t}(r \mid f_{k-1}) \leq 5 \), we set \( \alpha_{d} = \frac{3}{5} \alpha_{t} \).

Several parameters must be initialized or calculated prior to implementing the RCLMAP algorithm. For instance, the error covariance matrix \( \mathbf{E}^{-1} \) is initialized to an identity matrix multiplied by the variance of the previously estimated DVF. The term \( \gamma \) in (23) is a combination of the linearization error and noise power. The value for the variance of the linearization error is updated after each DVF is estimated, as was done in [19]. The noise power is estimated in a flat region of each image frame prior to processing, and the error in the intensity field estimates are also initialized to this value. The spatial gradients of \( f_{k-1} \) are required by the RCLMAP algorithm in (23) and (24). In order to evaluate these spatial gradients, we use a bilinear interpolation scheme described in (18). This technique is also adequate when moderate amounts of noise are present in the image sequence. However, to estimate the gradient in severe noise may require either a prefiltering or regularization step \cite{7}, \cite{29} to avoid noise amplification. Similarly, the accuracy of the spatial edges estimated by an edge detector is effected by the presence of noise. Therefore, due to its robustness to noise as well as its computational efficiency we use the edge detector described in (41).

To handle the raster scan transitions and boundary effects, we take an approach similar to the one presented in \cite{33}, where the effective size of the 2-D lattice is decreased by the boundary conditions. The result is that the DVF is not estimated in the region for which boundary conditions are
required. For a small region of support, such as the one used in this paper, the effect is not noticeable.

It is mentioned here that the computations of the proposed algorithm can be reduced in the following two ways: a) by using a steady-state Kalman gain instead of using (23) at each pixel location (it was found experimentally that the results obtained by a varying and a steady-state gain are almost identical) and b) by evaluating (26) not for all \( L \) possible configurations but instead for the most probable configurations of \( l(r) \) as identified by the energy function (9) and the knowledge of the state of the elements contained in \( \mathbb{R}^2 \) (dotted lines in Fig. 1). It was found experimentally that tracking the four most probable configurations produced results virtually identical to those obtained when all possible configurations are compared.

VI. EXPERIMENTAL RESULTS

In this section, we present several experimental results that illustrate the effectiveness of the RCLMAP algorithm and compare it with other widely used motion estimation algorithms. The algorithms were tested on a synthetic sequence and the 256 × 256, 8-bit real video sequences “Trevor White” and “Mobile.” A frame from the 128 × 128, 8-bit synthetic sequence “Phantom” is shown in Fig. 2. It is a sequence with a moving object in a moving background. In the sequence “Mobile,” an object moves across a differently moving background, which produces a DVF with several motion boundaries. The “Trevor White” sequence has an object moving against a stationary background; however, this is a noisy sequence with an estimated signal-to-noise ratio (SNR) of approximately 20 decibels (dB). For the purpose of assessing the performance of the RCLMAP algorithm, the Wiener-based PR (WPR) [4], BM [40], and extended Kalman filter (EKF) algorithms are applied to these sequences. The BM algorithm was implemented using 4 × 4 nonoverlapping blocks. To improve the quality of the estimated DVF, the values of the off-center vectors for each block were determined using bilinear interpolation. As mentioned in the previous section, the EKF is a special case of the RCLMAP algorithm and is implemented using (20)–(24) with the line process turned off (i.e., \( l(r) = [0, 0, 0, 0] \)).

The model parameters for the 16 VGM submodels, required by the RCLMAP algorithm, were determined as discussed in the previous section. The initial value for the variance of the linearization error \( \gamma \) was determined to be equal to 5. The covariance matrix \( \mathbf{E} \) was initialized to \( \sigma_{w,1,1,1}^2 \mathbf{I} \), where \( \sigma_{w,1,1,1}^2 \) is the variance of the noise term that corresponds to the model with all line elements activated (\( l(r) = [1, 1, 1, 1] \)).

The “Phantom” sequence was generated using the following scalar Gauss–Markov model

\[
\begin{align*}
    f'(m, n) &= 0.333 \cdot [f'(m, n - 1) + f'(m - 1, n)] \\
    &+ f'(m - 1, n - 1)] + n_i(m, n) 
\end{align*}
\]

where \( i = 1, 2 \), and \( n_i \) is a Gaussian random variable with mean \( \mu_i \) and variance \( \sigma_i^2 \). The background and the square object were created by setting \( \mu_1 = 50, \sigma_1^2 = 49 \) and \( \mu_2 = 100, \sigma_2^2 = 25 \), respectively. In each consecutive frame, the object was displaced, as compared to the previous frame, two pixels to the right against the background that was moving one pixel down and one pixel to the right. We applied all four algorithms to the “Phantom” sequence. In each case, only one iteration was used. Since the true motion in the “Phantom” sequence is known, we assess the accuracy of the estimated DVF using the mean squared error (MSE) and bias in the horizontal and vertical directions, which are defined as follows:

\[
\text{MSE}_i = \frac{1}{MN} \sum_r (d_i(r) - \hat{d}_i(r))^2, \quad (i = x, y) 
\]

\[
\text{bias}_i = \frac{1}{MN} \sum_r (d_i(r) - \hat{d}_i(r)), \quad (i = x, y). 
\]

The bias and MSE values for the DVF’s estimated using the four algorithms under consideration are shown in Table I. From these results, it is clear quantitatively that a higher degree of accuracy is obtained by the RCLMAP algorithm in estimating the DVF. The estimated DVF and corresponding line process obtained by the RCLMAP algorithm are shown in Figs. 3 and 4. These figures clearly demonstrate the RCLMAP algorithm’s ability to estimate, and thus preserve, the discontinuities present in the DVF. The cost parameters for the line process’ energy function were set as described in Section V, with \( \alpha_p = 10, \alpha_q = 10, \alpha_x = 13.3, \) and \( \alpha_y = 8 \).

The effects of additive noise (SNR’s of 25, 20, and 15 dB) on the accuracy of the estimated DVF for the synthetic
sequence is demonstrated in Table II, where the MSE and the bias of the estimate are shown again. In Figs. 5 and 6, the estimated DVF’s from the noisy “Phantom” sequence (SNR = 20 dB) obtained using the BM and RCLMAP algorithms, respectively, are presented. To reduce the effects of noise on the accuracy of the estimated DVF, it was necessary to increase the block size to 8 x 8 in the BM algorithm. From Fig. 5, it is evident that this increase in block size aggravates the blurring or over-smoothing of the motion boundaries. For the RCLMAP algorithm, the noise causes slight increases in the errors around the boundaries as well as the variance of the estimated DVF. The weights used to assign the relative importance of the terms in the line process remain unchanged, except for $\alpha_y$ (the weight associated with the assumption that line elements coincide with intensity edges), which is reduced. This is due to the decrease in reliability of estimated edge maps obtained from an edge detector operating in the presence of noise. For the SNR’s of 25, 20, and 15 dB, $\alpha_y$ was set to values of 8, 5, and 2, respectively, while $\gamma$ was increased in each case by the corresponding noise power.

For real image sequences, since the true DVF is not known, one way to evaluate the performance of a motion estimator is by using the improvement in motion compensation as a figure of merit. It is defined in decibels as

$$I_{mc}^f(k) = 10 \cdot \log_{10} \left( \frac{\sum_{k=2}^{K} \Delta^2 r_k \cdot d_k}{\sum_{k=2}^{K} \Delta^2 (r_k, d')} \right)$$

where $r_k$ denotes the spatial location $(m, n)$ on the $k$th frame and $I$ denotes the maximum number of iterations allowed in estimating $d$ at each location. Using all $K$ frames in the sequence, we define the average improvement in motion compensation in decibels as

$$I_{mc}^f = 10 \cdot \log_{10} \left( \frac{\sum_{k=1}^{K} \sum_{r_k} \Delta^2 (r_k, d')}{\sum_{k=2}^{K} \sum_{r_k} \Delta^2 (r_k, d')} \right).$$

A second metric for evaluating the accuracy of an estimated DVF is to consider the reconstruction of the original frames us-

---

**TABLE I**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Winter-based PR</th>
<th>BM</th>
<th>Extended Kalman Filter</th>
<th>RCLMAP $I = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MSE_r$</td>
<td>0.56</td>
<td>0.66</td>
<td>0.31</td>
<td>0.03</td>
</tr>
<tr>
<td>$MSE_s$</td>
<td>0.56</td>
<td>0.66</td>
<td>0.44</td>
<td>0.03</td>
</tr>
<tr>
<td>$bias_r$</td>
<td>0.17</td>
<td>0.13</td>
<td>0.21</td>
<td>0.04</td>
</tr>
<tr>
<td>$bias_s$</td>
<td>0.21</td>
<td>0.17</td>
<td>0.18</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>Winter-based PR</th>
<th>BM</th>
<th>Extended Kalman Filter</th>
<th>RCLMAP $I = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MSE_{E_k}$</td>
<td>20</td>
<td>0.67</td>
<td>0.7</td>
<td>0.31</td>
</tr>
<tr>
<td>$MSE_{E_s}$</td>
<td>25</td>
<td>0.63</td>
<td>0.83</td>
<td>0.50</td>
</tr>
<tr>
<td>$bias_{E_k}$</td>
<td>25</td>
<td>0.17</td>
<td>0.14</td>
<td>0.24</td>
</tr>
<tr>
<td>$bias_{E_s}$</td>
<td>25</td>
<td>0.71</td>
<td>0.17</td>
<td>0.5</td>
</tr>
<tr>
<td>$MSE_{E_k}$</td>
<td>20</td>
<td>0.70</td>
<td>0.74</td>
<td>0.51</td>
</tr>
<tr>
<td>$MSE_{E_s}$</td>
<td>20</td>
<td>0.83</td>
<td>1.0</td>
<td>0.55</td>
</tr>
<tr>
<td>$bias_{E_k}$</td>
<td>20</td>
<td>0.75</td>
<td>0.32</td>
<td>0.5</td>
</tr>
<tr>
<td>$bias_{E_s}$</td>
<td>20</td>
<td>0.89</td>
<td>0.35</td>
<td>0.49</td>
</tr>
<tr>
<td>$MSE_{E_k}$</td>
<td>15</td>
<td>0.76</td>
<td>0.99</td>
<td>0.71</td>
</tr>
<tr>
<td>$MSE_{E_s}$</td>
<td>15</td>
<td>0.95</td>
<td>0.96</td>
<td>0.81</td>
</tr>
<tr>
<td>$bias_{E_k}$</td>
<td>15</td>
<td>0.83</td>
<td>0.82</td>
<td>0.61</td>
</tr>
<tr>
<td>$bias_{E_s}$</td>
<td>15</td>
<td>0.96</td>
<td>0.34</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Fig. 7. $I_m^t (k)$ for the “Trevor White” sequence.

Fig. 8. PSNR(k) for the “Trevor White” sequence.

The improvement in performance obtained using the RCLMAP algorithm to estimate the DVF is further demonstrated in Figs. 9–12. The FD for frames 32 and 33 is shown in Fig. 9, while the DFD obtained using the RCLMAP algorithm ($I = 1$) is shown in Fig. 10. In order to display the values of both the FD and the DFD, the following mapping has been used

$$\Delta(r) = \max\{0, \min\{128 + \omega \cdot \Delta(r, r - d), 255\}\}$$

where $\omega$ is an approximate scale factor ($\omega = 6$ was used in this paper). Recall that if $d = 0$, then $\Delta$ represents the mapped FD. In comparing these figures, it is clear that the reduction in the prediction error obtained using the RCLMAP (Fig. 10) is considerable. This is especially evident near the boundaries of the moving object (Trevor). It is in these areas that the majority of the errors incurred by the three other algorithms were concentrated, demonstrating their inability to properly take into account the discontinuities present in the DVF. Based on the estimated line process, the RCLMAP algorithm adapts its prediction stage in order to preserve these discontinuities in the estimated DVF. The effectiveness of the RCLMAP algorithm at estimating both the line process and the DVF is shown in Figs. 11 and 12.

Similar results were obtained for frames 1 to 20 of the “Mobile” sequence. The average improvement $I_m^t$ and PSNR for the four algorithms under consideration are provided in Table III. From these results, it is again apparent that the RCLMAP algorithm significantly outperforms the other techniques. The same procedures discussed above for determining the parameters required by the RCLMAP algorithm were also used for this sequence (frames 5 and 6 were used for

Table III

<table>
<thead>
<tr>
<th>Estimation Algorithm</th>
<th>Wiener-based PR</th>
<th>BM</th>
<th>Extended Kalman Filter</th>
<th>RCLMAP $I = 1$</th>
<th>RCLMAP $I = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trevor $I_m^t$</td>
<td>4.44</td>
<td>4.5</td>
<td>6.96</td>
<td>10.15</td>
<td>12.50</td>
</tr>
<tr>
<td>PSNR</td>
<td>33.4</td>
<td>29.4</td>
<td>34.1</td>
<td>36.1</td>
<td>36.3</td>
</tr>
<tr>
<td>Mobile $I_m^t$</td>
<td>6.52</td>
<td>6.1</td>
<td>7.2</td>
<td>12.15</td>
<td>12.4</td>
</tr>
<tr>
<td>PSNR</td>
<td>27.6</td>
<td>24.1</td>
<td>28.9</td>
<td>31.9</td>
<td>32.0</td>
</tr>
</tbody>
</table>
calculating the prediction coefficients, while the remaining parameters were initialized to the same values as for the "Trevor White" sequence). The FD for frames 9 and 10 is shown in Fig. 13, and the DFD resulting by the use of the RCLMAP \((f = 1)\) algorithm is shown in Fig. 14. Closer inspection of these figures indicates that the DFD contains only a few degradations at the boundaries between the moving object and moving background. Figs. 15 and 16 show the estimated line process and corresponding DVF obtained for the above set of frames. Due to the amount of motion in the "Mobile" sequence, the estimated DVF was subsampled by a factor of four in both the horizontal and vertical directions, for display purposes. It is evident from these figures that both the estimated line process as well as the DVF correspond well to the actual motion in the sequence.

As a final comment, we would like to address the issue of the relative importance of the various constraints imposed on the line field through the energy function of (9). A detailed analysis of this issue is an extremely complicated problem [26]. In general, one has to study the effect of all possible combinations of six continuous variables ranging from 0 to \(\infty\). In an attempt to experimentally characterize the effectiveness of some of these constraints, we found that although all of them are important and necessary to achieve the demonstrated performance of the RCLMAP algorithm, their relative level of importance is dependent upon the motion content of the (frames in the) sequence. For instance, when determining the importance of the temporal continuity constraint (the term in
(9) weighted by $\alpha_v$, it was found that when this term was removed from the energy function, while the values of the rest of the weights were kept the same, the performance of the RCLMAP algorithm was degraded in parts of “Trevor White” by more than 1 dB, with respect to $J_{mc}(k)$. However, in other parts of the sequence, the degradation in performance was as little as 0.25 dB. This variation in the measured level of importance of a particular constraint is also due in part, to the strong interactions that exist between all the constraints. For example, the temporal continuity constraint interacts primarily with the intensity edge constraint (the term in (9) weighted by $\alpha_{eg}$).

VII. SUMMARY

In this paper, a vector coupled Gauss–Markov model for the DVF is proposed, and the corresponding MAP estimator is derived. The advantage of the proposed VCGM model is that the upper field provides the local characteristics for each displacement vector, while the line process is the mechanism for representing the structure of the DVF. It is the line process that provides the means for modeling the discontinuities associated with the DVF. Therefore, the spatially varying VCGM model provides a more realistic representation of the DVF than stationary models. A sophisticated model for the line
process was proposed, which accounts for all the desirable properties of DVF’s found in most dynamic sequences. In the formulation of the DVF estimator, we proposed employing the VCGM model with a causal neighborhood system to make it more attractive for real-time applications. Based on the VCGM model, we developed the computationally efficient reduced order model CLMAP estimator. The performance of the proposed RCLMAP algorithm was compared with the WPR, BM algorithms, and extended Kalman filter in estimating the DVF of synthetic and standard video image sequences. The performance of the RCLMAP algorithm was shown to be superior to that of the rest of the algorithms with respect to the accuracy of the estimated DVF, reduction of prediction and interpolation errors, and the ability to preserve the edges within the DVF. The developed algorithm was found to be robust to inaccuracies associated with the various required parameters. Therefore, these parameters do not need to be reoptimized for each set of frames. This helps in reducing the computational load associated with the RCLMAP algorithm.

Current research focuses on the problem of spatio-temporal filtering. Currently, many such algorithms [15, 30] rely on motion estimates obtained from a separate estimation process, after prefiltering. However, errors caused by the prefiltering can effect the accuracy of the DVF estimates. Also, separating the task of image sequence noise filtering from that of DVF estimation ignores the obvious relationship between the two. Therefore, we are currently developing algorithms that simultaneously estimate both the DVF as well as the intensity field from noisy or noisy blurred image sequences [8], [9].

REFERENCES


James C. Brailean (S'91-M'93) received the B.S.E.E. degree from the University of Michigan in 1985. He received the M.S.E.E. from the University of Southern California in 1988 and the Ph.D. in electrical engineering from Northwestern University in 1993.

From June 1985 to September 1989, he was a member of Technical Staff at Hughes Aircraft Co., Space and Communications Group, Los Angeles, CA, working on advanced communication systems. He is currently a Staff Engineer at Motorola, Schaumburg, IL, where he is working on video compression. His research interests are in video compression, motion estimation, and image sequence restoration.

Dr. Brailean is a Hughes Aircraft, Murphy, and Cabell Fellow. He is a member of Tau Beta Pi,Eta Kappa Nu, and SPIE.

Aggelos K. Katsaggelos (S'80-M'85-SM'92) received the Diploma degree in electrical and mechanical engineering from the Aristotle University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees, both in electrical engineering, from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, where he is currently an Associate Professor. During the 1986-87 academic year, he was an Assistant Professor at Polytechnic University, Department of Electrical Engineering and Computer Science, Brooklyn, NY. His current research interests include image recovery, processing of moving images (motion estimation, enhancement, very low bit rate compression), and computational vision.

Dr. Katsaggelos is an Ameritech Fellow and a member of the Associate Staff, Department of Medicine, at Evanston Hospital. He is a member of SPIE, the Steering Committees of the IEEE TRANSACTIONS ON MEDICAL IMAGING and the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE Technical Committees on Visual Signal Processing and Communications and on Image and Multi-Dimensional Signal Processing, the Technical Chamber of Commerce of Greece, and Sigma Xi. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990-92) and is currently an Area Editor for Graphical Models and Image Processing. He is the Editor of Digital Image Restoration (Springer-Verlag, Heidelberg, 1991) and the General Chairman of the 1994 Visual Communications and Image Processing Conference (Chicago, IL).