LIDS-P-2081

Sequential Filtering for Multi-Frame Visual Reconstruction *

Toshio M. Chin William C. Karl Alan S. Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute Of Technology
Cambridge, Massachusetts 02139

November 15, 1991

Abstract

We describe an extension of the single-frame visual reconstruction problem in which we consider how to efficiently and optimally fuse multiple frames of measurements obtained from images arriving sequentially over time. Specifically we extend the notion of spatial coherence constraints, used to regularize single-frame problems, to the time axis yielding temporal coherence constraints. An information form variant of the Kalman filter is presented which yields the optimal maximum likelihood estimate of the field at each time instant and is tailored to the visual field reconstruction problem. Propagation and even storage of the optimal information matrices for visual problems is prohibitive, however, since their size is on the order of $10^8 \times 10^8$ to $10^{12} \times 10^{12}$. To cope with this dimensionality problem a practical yet near-optimal filter is presented. The key to this solution is the observation that the information matrix, i.e. the inverse of the covariance matrix, of a vector of samples of a spatially distributed process may be precisely interpreted as specifying a Markov random field model for the estimation error process. This insight leads directly to the idea of obtaining low-order approximate models for the estimation error in a recursive filter through the recursive approximation of the information matrix by an appropriate sparse, spatially localized matrix. Numerical experiments are presented to demonstrate the efficacy of the proposed filter and approximations.

*This research was supported in part by the Office of Naval Research under Grant N00014-91-J-1004, the National Science Foundation under Grant MIP-9015281, and by the Army Research Office under Grant DAAL03-86-K-0171.
Address for correspondence: Alan Willsky, MIT, Room 35-437, Cambridge, MA 02139.

1 Introduction

Many low-level visual reconstruction problems are formulated as least squares problems with two types of constraint terms – constraints imposed by a static set of measurements obtained from the images and *smoothness* or *spatial coherence* constraints. Examples of these problems can be found in the computation of dense fields of depth [13, 14], shape [25, 26], and motion [24, 21]. Reconstructing such low-level visual fields from measurements made on a single image or a pair of images (as required, for example, in motion estimation) typically leads to under-constrained inverse problems, since we are trying to recover features (such as depth, shape, and motion) of objects in a 3-D domain from the *projected* information available in 2-D images [22]. The inclusion of spatial coherence constraints is by far the most common approach to regularizing these problems and ensuring the existence, uniqueness, and stability of the resulting solutions [2]. Such constraints take the form of cost terms penalizing the magnitudes of the spatial gradients of the unknown field. Physically, these cost terms correspond to the assumption that the sought after quantities have properties such as rigidity and smoothness [21].

Reconstructing visual fields by dynamically processing *sequences* of measurements has an obvious advantage over static reconstruction based on a single data set. For one thing, the accumulation of a larger quantity of data leads to a more reliable estimate due to a reduction in measurement noise. Another advantage, not as obvious, is that in some cases a single frame of data may not provide sufficient information to resolve static ambiguities, and hence for reasonable estimates to be obtained, temporal information must be utilized as well. For example, in optical flow estimation we wish to estimate a two-dimensional motion vector at each pixel location using one-dimensional measurements of intensity changes at each pixel. The use of spatial coherence constraints makes it

possible to resolve the ambiguity in the problem as long as the intensity field has substantial spatial diversity in the direction of its spatial gradient [23]. However, if the measured spatial gradients have identical (or nearly identical) directions over the entire image frame, any motion perpendicular to this spatial gradient is unresolvable (or highly uncertain) from a single data frame [24, 21]. On the other hand, in many cases the desired diversity of gradient directions *is* available over *time*, allowing the resolution of this ambiguity by incorporating more frames of measurements [6].

In this paper, we describe an extension of the classical single-frame reconstruction problem in which we consider fusing multiple frames of measurements obtained from images arriving sequentially over time. Specifically we examine the straightforward extension of the classical spatial coherence constraints to the time axis yielding *temporal coherence constraints* [15]. Such constraints are represented by the addition of cost terms involving temporal derivatives. We formulate these multi-frame visual field reconstruction problems in an estimation-theoretic framework. The single-frame problem can be formulated as an estimation problem [34, 37], so that the computed visual field can be considered as a jointly Gaussian random field. By capturing the time evolution of the field probabilistically this formulation allows us to treat a sequence of unknown fields $\underline{\mathbf{f}}(t)$ indexed by time t as a space-time stochastic process. Conceptually, we can then utilize well-developed optimal sequential estimation algorithms, such as the Kalman filter and its derivatives.

Unfortunately, for typical image-based applications the dimension of the associated state will be on the order of the number, N, of pixels in the image, typically 10^4 to 10^6 elements. The associated covariance matrices for an optimal filter are thus on the order of $10^8 \times 10^8$ to $10^{12} \times 10^{12}$! The storage and manipulation of such large matrices, as required by the optimal filters, is clearly prohibitive, necessitating the use of a sub-optimal approach. In the past [19, 16, 17, 18, 31, 37] ad hoc methods have been used to obtain computationally feasible algorithms. In contrast, in this

paper we examine in detail the structure of the optimal filter and pinpoint both the source of its computational complexity and the route to the systematic design of nearly optimal approximations. In particular, the key to our approach is the observation that the information matrix, i.e. the inverse of the covariance, of the estimation error in the Kalman filter estimate, has a natural interpretation as a Markov random field model for the estimation error. This suggests the idea of seeking sparse, spatially-local, low-order approximations to such models which in essence make each stage of the recursive estimation algorithm no more complex than the solution to static visual field reconstruction algorithms. Such approximations will be shown to yield near-optimal results, reflecting the fact that visual fields and the associated spatial coherence constraints are dominated by inherently local interactions. Our model-based approach provides a rational basis for computationally feasible, nearly optimal filter design for visual field reconstruction which naturally incorporates both temporal and spatial coherence constraints. The value of our approximations are demonstrated through numerical experiments.

2 Coherence Constraints and Maximum Likelihood Estimation

2.1 Coherence Constraints

Spatial Coherence: The Single-Frame Problem

We consider the problem of reconstructing a visual field $f(\underline{s})$ over a K-dimensional spatial domain $\mathcal{D}$ ($\underline{s} \in \mathcal{D} \subset \Re^K$) based on measurements $g(\underline{s})$ and $h(\underline{s})$ from a sequential set of images. The standard way in which such single-frame visual field reconstruction problems are formulated is given by [23]:

$$\min_{f(\underline{s})} \int_{\mathcal{D}} \nu(\underline{s}) \left\| g(\underline{s}) - h(\underline{s}) f(\underline{s}) \right\|^2 + \sum_i \mu_i(\underline{s}) \left\| \frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} f(\underline{s}) \right\|^2 d\underline{s} \qquad (1)$$

where $\nu(\underline{s}) \neq 0$ and $\mu_i(\underline{s})$ are strictly positive weighting parameters. We denote the components of the spatial index vector $\underline{s}$ by $s_k, k = 1, 2, \ldots, K$. The dimension K of the spatial domain in most visual reconstruction problems is at most 3. The orders i_k of the partial derivatives are non-negative integers, and $\frac{\partial^0}{\partial s_k{}^0} \equiv 1$. The index i, where $i = 1, 2, \ldots$, is used to distinguish the K-tuples $(i_1, i_2, \ldots, i_K)$. In reconstruction of a *vector* visual field, $f(\underline{s})$ and $g(\underline{s})$ become vector functions of $\underline{s}$ while $h(\underline{s})$ is a matrix function. An example of such a problem is found in the case of optical flow reconstruction [23], where $f(\underline{s})$, $g(\underline{s})$, and $h(\underline{s})$ have respective dimensions of 2×1, 1×1, and 1×2.

The first integrand term in (1) constrains the unknown field $f(\underline{s})$ based on the measurements $g(\underline{s})$ and $h(\underline{s})$. The spatial coherence constraint is expressed in (1) as the sum of quadratic terms involving spatial derivatives of the unknown field $f(\underline{s})$. First and second order derivatives are most commonly used. While spatial coherence constraints make the reconstruction problems mathematically well-posed by supplementing the measurement constraints [2], they can also be considered to be our prior knowledge about the unknown field before measurements are made [4, 37]. Such commonly used prior models include first-order differential constraints corresponding to a membrane model [23, 24, 21], second order differential constraints corresponding to a thin-plate model [13, 14, 18, 19] and hybrid constraints combining both first and second order derivatives to model the structure of object boundary contours [28].

Extension to Temporal Coherence

We now consider the straightforward extension of spatial coherence constraints over the time axis yielding *temporal coherence constraints.* Such constraints consist of cost terms involving temporal derivatives. Specifically, consider the following temporal extension of the general single-frame visual

reconstruction problem (1):

$$\min_{f(\underline{s},t)} \int_0^T \int_{\mathcal{D}} \nu(t) \left\| g(\underline{s},t) - h(\underline{s},t) f(\underline{s},t) \right\|^2 + \sum_i \mu_i(\underline{s},t) \left\| \frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} f(\underline{s},t) \right\|^2 \quad (2)$$
$$+ \sum_{i,j} \rho_{ij}(\underline{s},t) \left\| \frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} \frac{\partial^j}{\partial t^j} f(\underline{s},t) \right\|^2 d\underline{s}\, dt$$

where $f(\underline{s},t)$, $g(\underline{s},t)$, $h(\underline{s},t)$, $\nu(\underline{s},t)$, $\mu_i(\underline{s},t)$, and $\rho_{ij}(\underline{s},t)$ are now space-time functions. Note that the full solution to the optimization problem (2) leads to a reconstructed space-time field $\hat{f}(\underline{s},t)$, $\underline{s} \in \mathcal{D}$, $0 \leq t \leq T$ in which the reconstruction at any time takes advantage of all available constraints over the entire time interval. In the parlance of estimation theory, this is the optimal, noncausal, *smoothed* estimate. In this paper we focus on the optimal causal estimate, i.e. the value of the solution to (2) at the current time $t = T$. Thus as T increases we in fact are solving a ***different*** optimization problem for each T. As we will see, by adopting an estimation-theoretic perspective we can use Kalman-filter-like algorithms to perform such a calculation recursively. Furthermore, if the noncausal smoothed estimates are desired, one can use standard two-filter solutions, obtained by combining a causally (Kalman) filtered estimate with an anticausally filtered estimate. Thus the algorithms described herein also form the basis for the solution of the full optimization problem (2).

2.2 A Maximum Likelihood Formulation

We now describe an estimation theoretic formulation of the general low-level visual field reconstruction problem (2). We *interpret* the resulting least-squares formulation as an estimation problem, facilitating the development of efficient multi-frame reconstruction algorithms as presented in the next section. Bayesian estimation perspectives on visual reconstruction problems have been intro-

duced before [10, 34, 37, 18]. Casting these problems into a strictly Bayesian framework is somewhat awkward, however, because in many cases the variables to be estimated do not have well-defined probability densities. The maximum likelihood (ML) estimation framework described here provides us with a more natural way to express the reconstruction problem in an estimation-theoretic context by viewing spatial coherence as a noisy observation and temporal coherence as a dynamic equation driven by white noise. This mapping leads to a general dynamic representation for stochastic processes modeling various types of potential temporal behavior of visual fields. Such an approach provides a basis for *rational* integration of multiple frames of data in the estimation process, with its associated advantages of noise and ambiguity reduction. Moreover, the ML framework lends itself nicely to the use of *descriptor dynamic systems* [32, 33] for multi-frame reconstruction, which permit a wider range of temporal dynamic representations for the fields than are possible with the traditional Gauss-Markovian state-space systems and which deal in a convenient way with random and unknown variables, whether they do or do not have well-defined prior densities.

Let us write $x \sim (\, m, P \,)$ to denote that x is a Gaussian random vector with mean m and covariance P. We call the vector-matrix pair $(\, m, P \,)$ the *mean-covariance pair* associated with x. With this notation, the minimizing $f(\underline{s}, T)$ for the variational problem (2) may be found as the ML estimate for $f(\underline{s}, T)$ based on the set of dynamic equations:

$$\frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} \frac{\partial^{j}}{\partial t^{j}} f(\underline{s}, t) \;=\; q_{ij}(\underline{s}, t), \qquad q_{ij}(\underline{s}, t) \sim \Big(0, \rho_{ij}^{-1}(\underline{s}, t) \Big) \tag{3}$$

coupled with the set of observation equations:

$$g(\underline{s}, t) \;=\; h(\underline{s}, t) f(\underline{s}, t) + r_0(\underline{s}, t), \qquad r_0(\underline{s}, t) \sim \Big(0, \nu^{-1}(\underline{s}, t) \Big) \tag{4}$$

$$0 = \frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} f(\underline{\mathbf{s}}, t) + r_i(\underline{\mathbf{s}}, t) \quad r_i(\underline{\mathbf{s}}, t) \sim \left(0, \mu_i^{-1}(\underline{\mathbf{s}}, t) \right) \tag{5}$$

for $0 \leq t \leq T$ where the range of the subscripts i and j are the same as in the summations (2). Thus the $q_{ij}(\underline{\mathbf{s}}, t)$ and $r_i(\underline{\mathbf{s}}, t)$ along with $r_0(\underline{\mathbf{s}}, t)$ are *zero-mean* space-time white noise processes. The set of equations (3), (4), and (5) formulates (2) as an equivalent sequential ML estimation problem.

Viewing the spatial coherence constraints as a set of observations and the temporal coherence constraints as a set of dynamic equations for the field is useful in gaining insights into how pieces of information about the unknown $f(\underline{\mathbf{s}}, t)$ are each represented. For example, (4) represents the contribution from the measurements in the images, and (5) represents the prior spatial knowledge of the field as provided by the spatial coherence constraints. In particular, the prior model implied by the spatial coherence constraints is expressed in the ML estimation framework as a set of observations (5), each indicating that the differential $\frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} \cdots \frac{\partial^{i_K}}{\partial {s_K}^{i_K}} f(\underline{\mathbf{s}}, t)$ of the unknown is observed to be zero, the ideal situation, with an uncertainty of variance $\mu_i^{-1}(\underline{\mathbf{s}}, t)$. Similarly, the prior temporal model implied by the temporal coherence constraints is now expressed as the explicit set of dynamic equations (3) driven by white noise processes.

In practice, our measurements are only available over a *discrete* spatial and temporal domain. Rather than using the continuous formulation represented by (3), (4), and (5), we focus from this point forward on a discrete, vectorized formulation representing a discrete counterpart to these equations, using a regular 2-D sampling grid. First we treat discreteization of the observation equations (4) and (5) then the dynamic equations (3). For clarity of presentation we will also assume for the rest of the paper that the field is defined over a 2-D space (i.e. $K = 2$) and that the field is scalar. The filtering techniques to be presented in the sequel can be straightforwardly

extended to other cases such as vector fields and fields defined over a 1-D or 3-D space, as detailed in [6].

Observations

Define $\underline{\mathbf{f}}(t)$ to be a column vector whose components are $f(s_k, t)$, where $\{s_k\}$ denotes the set of points in the regular 2-D sample grid, ordered lexicographically. The measurement vector $\underline{\mathbf{g}}(t)$ is similarly defined from $g(s_k, t)$. We define the diagonal matrix $\mathbf{H}(t)$ to be one whose diagonal is formed from the elements of $h(s_k, t)$, in a matching lexicographic order, and the diagonal matrix $\mathbf{N}(t)$ to be one whose diagonal components are given by $\nu(s_k, t)$, similarly ordered. With these definitions we may represent the sampled version of the observations (4) in the form:

$$\underline{\mathbf{g}}(t) = \mathbf{H}(t)\underline{\mathbf{f}}(t) + \underline{\mathbf{r}}_0(t), \quad \underline{\mathbf{r}}_0(t) \sim \left(0, \mathbf{N}^{-1}(t) \right) \tag{6}$$

Next, let us examine the discrete counterpart of (5). Specifically, we define the matrix operator $\mathbf{S}_i$ to be a finite difference approximation to the cross differentiation operation so that

$$\mathbf{S}_i \, \underline{\mathbf{f}}(t) \approx \left[\frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}} f(\underline{\mathbf{s}}, t) \right]$$

where, as before, the index i is used to distinguish the K-tuples (i_1, i_2), for K=2. With this definition, (5) may be written in a discrete setting as:

$$0 = \mathbf{S}_i \, \underline{\mathbf{f}}(t) + \underline{\mathbf{r}}_i(t), \quad \underline{\mathbf{r}}_i(t) \sim \left(0, \mathbf{M}_i^{-1}(t) \right), \quad i = 1, 2, \ldots \tag{7}$$

where $\mathbf{M}_i(t)$ is a diagonal matrix whose diagonal is formed from the elements of $\mu_i(s_k, t)$ in a match-

ing lexicographic order. Equations (6) and (7) may be combined into the composite observation equation:

$$\underline{\mathbf{y}}(t) = \mathbf{C}(t)\underline{\mathbf{f}}(t) + \underline{\mathbf{r}}(t), \quad \underline{\mathbf{r}}(t) \sim (\ 0, \mathbf{R}(t)\ ) \tag{8}$$

where the component matrices are defined as follows:

$$\underline{\mathbf{y}}(t) \equiv \begin{bmatrix} \underline{\mathbf{g}}(t) \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad \mathbf{C}(t) \equiv \begin{bmatrix} \mathbf{H}(t) \\ \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \end{bmatrix}, \quad \mathbf{R}(t) \equiv \begin{bmatrix} \mathbf{N}^{-1}(t) & & & \\ & \mathbf{M}_1^{-1} & & \\ & & \mathbf{M}_2^{-1} & \\ & & & \ddots \end{bmatrix}. \tag{9}$$

Structure of $\mathbf{S}_i$

The matrices $\mathbf{S}_i$ have a special sparse and banded structure reflecting the fact that they represent finite difference operators. In particular, let $\mathbf{S}^{(i_1,i_2)}$ denote the matrix difference operators for a discrete, rectangular domain of size $n_1 \times n_2$ which correspond to the differentials $\frac{\partial^{i_1}}{\partial {s_1}^{i_1}} \frac{\partial^{i_2}}{\partial {s_2}^{i_2}}$. Then, the first-order difference operators are given by

$$\mathbf{S}^{(0,1)} = \begin{bmatrix} \Delta^{(1)} & & \\ & \ddots & \\ & & \Delta^{(1)} \end{bmatrix}, \quad \mathbf{S}^{(1,0)} = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \\ & \ddots & \ddots \\ & & -\mathbf{I} \quad \mathbf{I} \end{bmatrix},$$

while the second order difference operators are given by

$$\mathbf{S}^{(0,2)} = \begin{bmatrix} \Delta^{(2)} & & \\ & \ddots & \\ & & \Delta^{(2)} \end{bmatrix}, \quad \mathbf{S}^{(2,0)} = \begin{bmatrix} \mathbf{I} & -2\mathbf{I} & \mathbf{I} & & \\ & \ddots & \ddots & \ddots & \\ & & \mathbf{I} & -2\mathbf{I} & \mathbf{I} \end{bmatrix},$$

$$\mathbf{S}^{(1,1)} = \begin{bmatrix} -\Delta^{(1)} & \Delta^{(1)} & & \\ & \ddots & \ddots & \\ & & -\Delta^{(1)} & \Delta^{(1)} \end{bmatrix},$$

where the identity matrices $\mathbf{I}$ have dimension $n_1 \times n_1$ and Δ's are matrices of matching dimension such that

$$\Delta^{(1)} \equiv \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad \Delta^{(2)} \equiv \begin{bmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix}.$$

Dynamic Equations

Now we treat discretization of the dynamic equations (3). For clarity we only examine the case of *first-order* temporal derivative constraints, corresponding to $j = 1$ in (3), though we allow cross space-time constraints. The more general case involving higher-order temporal derivative constraints is treated in [6].

We may approximate the first-order temporal derivative in (3) by a first difference and use the discrete approximations $\mathbf{S}_i$ to the spatial derivatives of the previous section to obtain the following

general first-order temporal dynamic model for the field $\underline{\mathbf{f}}(t)$:

$$\mathbf{B}\underline{\mathbf{f}}(t) = \mathbf{B}\mathbf{A}(t)\underline{\mathbf{f}}(t-1) + \underline{\mathbf{q}}(t), \quad \underline{\mathbf{q}}(t) \sim (\ 0, \mathbf{Q}(t)\ ) \tag{10}$$

where, to correspond to (3), we need $\mathbf{B} = [\mathbf{S}_1^T \cdots \mathbf{S}_m^T]^T$ (capturing the cross space-time derivative constraints), $\mathbf{A}(t) = \mathbf{I}$, and $\mathbf{Q}(t)$ to be a diagonal matrix whose diagonal entries are the appropriately ordered $\rho_{ij}^{-1}(s_k, t)$.

Of course, given models of the form (10), we may make other choices for the system matrices $\mathbf{B}$, $\mathbf{A}(t)$, and $\mathbf{Q}(t)$ than those strictly corresponding to the continuous formulation (3). For example, the matrix $\mathbf{A}(t)$ allows the possibility of time-varying system dynamics. In some reconstruction problems the system matrix $\mathbf{A}(t)$ may play the important role of registering the moving visual field onto the image frame – a fundamental issue in multi-frame visual reconstruction. In such cases, the matrix $\mathbf{A}(t)$ provides the system model with the information of how the estimated field from time $t-1$ should be warped in order to fit into the image frame at time t, essentially performing local position adjustments of the components of the field via shifting and averaging [36, 18]. The matrix usually has a sparse structure in which the non-zero elements are concentrated around the main diagonal.

The model given in (10) is in the standard *descriptor* form [32, 33]. This model becomes a standard Gauss-Markov model if no spatial coherence constraints are applied to the temporal variation so that $\mathbf{B} = \mathbf{I}$. Thus the descriptor form naturally captures cross space-time differential constraints.

Overall Model

Combining the observation model (8) with the dynamic model given in (10) yields the following discrete system model we will use for estimation purposes:

$$\mathbf{B}\underline{\mathbf{f}}(t) = \mathbf{B}\mathbf{A}(t)\underline{\mathbf{f}}(t-1) + \underline{\mathbf{q}}(t), \quad \underline{\mathbf{q}}(t) \sim (\, 0, \mathbf{Q}(t)\,) \tag{11}$$

$$\underline{\mathbf{y}}(t) = \mathbf{C}(t)\underline{\mathbf{f}}(t) + \underline{\mathbf{r}}(t), \quad \underline{\mathbf{r}}(t) \sim (\, 0, \mathbf{R}(t)\,), \tag{12}$$

where the Gaussian noise processes $\underline{\mathbf{q}}(t)$ and $\underline{\mathbf{r}}(t)$ are uncorrelated over time. Note that the descriptor dynamic model (11) can be reformulated as a standard Gauss-Markov model if the matrix $\mathbf{B}$ has full column rank. In general, however, the resulting system model will not retain the nice sparse structure usually present in (10), making the present form preferred.

3 Sequential ML Estimation

In this section we examine sequential ML estimation of the field $\underline{\mathbf{f}}(t)$ given the model specified by (11) and (12). In solving such an ML estimation problem, one is ultimately interested in obtaining the posterior mean-covariance pair. A well-known solution to many recursive estimation problems of this type is the Kalman filter which provides a recursive procedure for propagating the desired mean-covariance pair. In its standard form, the Kalman filter represents the solution to a *Bayesian* estimation problem in which *prior* information, corresponding to probabilistic specification of initial conditions, and noisy dynamics are combined with real-time measurements to obtain conditional statistics. For such a method to make strict sense, the prior information must be sufficient to imply a well-defined prior distribution for the variables to be estimated. For the problems of

interest here, this is *not* the case in general and, in fact, is essentially never the case for regularized problems arising in computer vision. In particular, note that the matrices $\mathbf{S}_i$, corresponding to differential operators, are certainly not invertible, and hence $\mathbf{B}$ in (11) is typically singular (and often non-square). Thus (11) does *not* provide a prior distribution for $\underline{\mathbf{f}}(t)$.

For this reason, as well as several others, we adopt an alternate, implicit representation for the mean-covariance pair, called the *information pair* in which we propagate the *information matrix*, i.e. the inverse of the covariance matrix. As we will see, the information pair provides several advantages for the recursive estimation problems of interest here. The first is that this pair is always well-defined for our problems. In particular, what characterizes all regularization problems arising in computer vision is that, while neither the measurements nor smoothness constraints have unique minimizers individually, their joint minimization, as in (2), does have such a unique solution. In the context of our estimation problem, typically neither the dynamics (11) nor the measurements (12) separately provide full probabilistic information about $\underline{\mathbf{f}}(t)$, but together they do. Since the information filter form directly propagates and fuses information, whether in the form of noisy dynamic constraints (11) or noisy measurements (12), it is well-suited to these problems.

There are several other reasons that the information pair is of considerable interest. First, the fusing of statistical data contained in independent observations corresponds to a simple sum of the corresponding information matrices yielding computationally simple algorithms. Secondly, for problems of interest to us, in which the measurements (12) are local, the resulting information matrices, while not being strictly banded, are almost so, and thus can be well-approximated by sparse banded matrices. Such approximations provide us with a convenient and firm mathematical foundation on which we can design computational algorithms for visual reconstruction while reducing both the computational and storage requirements of any implementation. In particular,

since in most vision problems the observation matrix $\mathbf{C}(t)$ is in fact data-dependent, the error covariance matrix and gain, or their information pair equivalents, must not only be stored but also calculated on-line, making the issue of computational efficiency even more severe. Indeed as we will see, the measurement update step, i.e. when we incorporate the next measurement (12), involves adding a sparse, diagonally-banded matrix to the previous information matrix, enhancing diagonal dominance and in fact preserving banded structure if such structure existed before the update. The prediction step, i.e. when we use (11) to predict ahead one time step, does not strictly preserve this structure, and it is this point that dramatically increases the computational complexity of the fully optimal algorithm and which suggests the approximations developed in this paper. In particular, as we will see, the information matrix has the interpretation of specifying a Markov random field (MRF) model for the corresponding estimation error, and our approximation has a natural interpretation as specifying a reduced-order local MRF model for this error.

3.1 Information Form of ML Estimate

To this end, consider a general ML estimation problem for an unknown $\underline{\mathbf{f}}$ based on the observation equation $\underline{\tilde{\mathbf{y}}} = \tilde{\mathbf{C}}\underline{\mathbf{f}} + \underline{\tilde{\mathbf{r}}}$, $\underline{\tilde{\mathbf{r}}} \sim \left(0, \tilde{\mathbf{R}}\right)$. We call the quantities $\underline{\mathbf{z}} \equiv \tilde{\mathbf{C}}^T\tilde{\mathbf{R}}^{-1}\underline{\tilde{\mathbf{y}}}$ and $\mathbf{L} \equiv \tilde{\mathbf{C}}^T\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{C}}$ the *information pair* associated with the unknown $\underline{\mathbf{f}}$. We use double angular brackets as in $\underline{\mathbf{f}} \sim \langle\langle\, \underline{\mathbf{z}}, \mathbf{L} \,\rangle\rangle$ to denote information pairs in order to distinguish them notationally from mean-covariance pairs. The matrix $\mathbf{L}$ is just the *information matrix* or *observation grammian* of the problem [27].

In the visual reconstruction problems considered here, $\mathbf{L}$ is always invertible when the estimates are based on both the measurement and coherence constraints as in (8). The information matrix $\mathbf{L}$ tends not be invertible, however, when one attempts to solve the problems based only on the measurements as in (6) (corresponding to an ill-posed formulation) or to obtain Bayesian priors

based only on the coherence constraints as in (7).

In the case that $\mathbf{L}$ is invertible, the estimate $\hat{\underline{\mathbf{f}}}$ and error covariance $\mathbf{P}$ for the ML estimation problem can be obtained from the corresponding information pair as:

$$\hat{\underline{\mathbf{f}}} = \mathbf{L}^{-1}\underline{\mathbf{z}} \tag{13}$$

$$\mathbf{P} = \mathbf{L}^{-1}. \tag{14}$$

Thus, the information pair $\langle\langle\, \underline{\mathbf{z}}, \mathbf{L} \,\rangle\rangle$ contains the same statistical data as those in the mean-covariance pair $\left(\hat{\underline{\mathbf{f}}}, \mathbf{P}\right)$. Specifically, the information pair expresses the solution of the ML estimation problem *implicitly* in the sense that the estimate is given as the solution of the inverse problem:

$$\mathbf{L}\hat{\underline{\mathbf{f}}} = \underline{\mathbf{z}}. \tag{15}$$

An important point to note is that in image processing problems the vector $\underline{\mathbf{f}}$ is of extremely high dimension so that calculating $\hat{\underline{\mathbf{f}}}$ by direct computation of $\mathbf{L}^{-1}$ as in (13) is prohibitive. However, if $\mathbf{L}$ is a sparse, banded matrix, as it is, for example, in single-frame computer vision problems, then (15) may be solved more efficiently using, for example, Gauss-Seidel iterations [24] or multigrid methods [38, 39].

3.2 An Information Based Filter

In this section we present an optimal information filtering algorithm for the system (11), (12) which is a variant of the information form of the Kalman filter [3]. A detailed derivation may be found in [6]. Let $\mathbf{U}(t) \equiv \mathbf{B}^T\mathbf{Q}^{-1}(t)\mathbf{B}$, then the optimal ML estimate and its corresponding information

pair are obtained from the recursive algorithm:

- Prediction:

$$\bar{\mathbf{L}}(t) = \mathbf{U}(t) - \mathbf{U}(t)\mathbf{A}(t)\left(\mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \hat{\mathbf{L}}(t-1)\right)^{-1}\mathbf{A}^T(t)\mathbf{U}(t) \tag{16}$$

$$\bar{\underline{\mathbf{f}}}(t) = \mathbf{A}(t)\hat{\underline{\mathbf{f}}}(t-1) \tag{17}$$

$$\bar{\underline{\mathbf{z}}}(t) = \bar{\mathbf{L}}(t)\bar{\underline{\mathbf{f}}}(t) \tag{18}$$

- Update:

$$\hat{\mathbf{L}}(t) = \bar{\mathbf{L}}(t) + \mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t) \tag{19}$$

$$\hat{\underline{\mathbf{z}}}(t) = \bar{\underline{\mathbf{z}}}(t) + \mathbf{C}^T(t)\mathbf{R}^{-1}(t)\underline{\mathbf{y}}(t) \tag{20}$$

$$\hat{\mathbf{L}}(t)\hat{\underline{\mathbf{f}}}(t) = \hat{\underline{\mathbf{z}}}(t). \tag{21}$$

When the descriptor dynamic equation (11) can be expressed in a Gauss-Markov form, standard information Kalman filtering equations (e.g., [1, 30]) can also be applied to the estimation problem. The filtering algorithm (16)–(21) is, however, more suitable for visual reconstruction mainly due to the fact that, unlike traditional information Kalman filters, the inverse of $\mathbf{A}(t)$ is not needed. As previously mentioned, in visual reconstruction the system matrix $\mathbf{A}(t)$ often performs a local averaging and thus is sparse. Taking its inverse generally loses its sparseness and thus the associated computational efficiency of the filter. It is also conceivable that $\mathbf{A}(t)$ may not even be invertible in some formulations.

Let us close this section with several observations that serve to motivate and interpret the development in the next section. First, note that the matrix $\mathbf{C}(t)$ constructed in Section 2 for regularized

computer vision problems is composed of sparse and banded blocks and $\mathbf{R}(t)$ is diagonal so that $\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)$ itself is sparse and banded. Thus if $\overline{\mathbf{L}}(t)$ is also sparse and similarly banded, then so is $\widehat{\mathbf{L}}(t)$, making inversion of (21) computationally feasible. However, while information matrices, i.e. inverses of covariances, add in the update step, it is covariances that add in the prediction step[†]. When this addition is represented in terms of information matrices, as in (16), we find that the banded structure of $\mathbf{L}$ is *not* preserved by the prediction step, since the inverse of the matrix $\mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \widehat{\mathbf{L}}(t-1)$ will *not* be banded in general even if $\widehat{\mathbf{L}}(t-1)$ is banded.

While the exact implementation of (16)–(21) involves full matrices, there are strong motivations for believing that nearly optimal performance can be achieved with banded approximations. Note first that the banded structure of $\mathbf{C}(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)$ is such that if $\overline{\mathbf{L}}(t)$ is diagonally dominant or, more generally, has its significantly nonzero values in a sparse band, then the summation in (19) will enhance this property in $\widehat{\mathbf{L}}(t)$. This property in turn implies, through (16) evaluated at $(t+1)$, that $\overline{\mathbf{L}}(t+1)$ will also inherit this property. This observation suggests the idea of developing recursive procedures involving banded approximations to the information matrices, and this is the approach pursued in the next section.

A useful way in which to interpret such an approximation is given by a closer examination of (19)–(21), in which we are fusing previous information, as captured by $\left\langle\left\langle\, \overline{\underline{\mathbf{z}}}, \overline{\mathbf{L}} \,\right\rangle\right\rangle$, with the new data $\underline{\mathbf{y}}(t)$. This new data vector is used in essence to estimate (and thus reduce) the error in the estimate $\overline{\underline{\mathbf{f}}}(t)$ prior to update. In particular, let this error be given by $\tilde{\underline{\mathbf{f}}} = \underline{\mathbf{f}} - \overline{\underline{\mathbf{f}}}$ and define:

$$\tilde{\underline{\mathbf{y}}}(t) = \underline{\mathbf{y}}(t) - \mathbf{C}(t)\overline{\underline{\mathbf{f}}}(t) = \mathbf{C}(t)\tilde{\underline{\mathbf{f}}}(t) + \underline{\mathbf{r}}(t) \tag{22}$$

[†]This is just a generalization of the fact that the covariance of the sum of independent random vectors is the sum of their covariances

If $\hat{\tilde{\underline{f}}}(t)$ denotes the best estimate of $\tilde{\underline{f}}(t)$ based on $\tilde{\underline{y}}(t)$ and its prior information pair $\left\langle\left\langle\, 0, \overline{\mathbf{L}}(t) \,\right\rangle\right\rangle$, then $\hat{\underline{f}}(t)$ in (21) exactly equals $\tilde{\underline{f}}(t) + \hat{\tilde{\underline{f}}}(t)$. Thus, the update step (19)–(21) is nothing more than a static estimation problem for $\tilde{\underline{f}}(t)$. Furthermore, the information pair $(0, \overline{\mathbf{L}}(t))$ for $\tilde{\underline{f}}(t)$ can very naturally be thought of as a *model* for $\tilde{\underline{f}}(t)$ of the following form:

$$\overline{\mathbf{L}}(t)\tilde{\underline{f}}(t) = \underline{\zeta}(t) \tag{23}$$

where $\underline{\zeta}(t)$ is zero mean and has covariance $\overline{\mathbf{L}}(t)$ (so that the covariance of $\tilde{\underline{f}}(t)$ is $\overline{\mathbf{L}}^{-1}(t)$). Furthermore, the model (23) corresponds to an MRF model for $\tilde{\underline{f}}(t)$ with a neighborhood structure determined by the locations of the nonzero elements of $\overline{\mathbf{L}}(t)$. For example, in the case of 1-D MRF's, a tridiagonal $\overline{\mathbf{L}}(t)$ corresponds exactly to a 1-D nearest neighbor MRF [29, 12, 41], and analogous banded structures, described in the next section, correspond to nearest-neighbor and other more fully-connected MRF models in 2-D. From this perspective we see that seeking banded approximations to information matrices corresponds in essence to *reduced-order MRF modeling* of the error in our spatial estimates at each point in time.

Interpreting the information matrix $\mathbf{L}$ as an MRF model for the estimation error $\tilde{\underline{f}}$ can be quite useful, as it connects our ML estimation formulations directly with other important formulations in visual field reconstruction, such as detection of discontinuities [10, 20], and provides a rational basis for the design of sub-optimal filters, as discussed in the following section.

4 A Sub-Optimal Information Filter

For typical image-based applications the dimension of the state in the corresponding model (11) will be on the order of the number of pixels N in the image data, typically 10^4 to 10^6 elements. The

associated information matrices for the optimal filter of Section 3 are $O(N^2)$ so that implementation of these optimal filters would require the storage and manipulation of $10^8 \times 10^8$ to $10^{12} \times 10^{12}$ matrices! As a result, practical considerations require some sort of approximate, sub-optimal scheme.

Indeed, our optimal algorithm has been designed from the outset to minimize the number of approximations necessary for implementation. In this section we show how to approximate the optimal filter in a rational way that retains nearly optimal behavior. We want our approximations to achieve 1) a reduction in the storage requirements for the required information matrices and 2) a reduction in the on-line computational burden of the filters (particularly as imposed by any matrix inverses or factorizations). Where possible we also seek to achieve enhanced parallelizability of the algorithm. We effectively achieve these goals by approximating the multi-frame algorithm so that the resulting sub-optimal filters are truly local and thus parallelizable and require much reduced memory for matrix storage. The key to these goals lies in exploiting and preserving sparseness of the information matrices $\mathbf{L}$. Alternatively, as we have pointed out, these approximations will be seen to be equivalent to the identification of a reduced order model of fixed and specified structure at each prediction step.

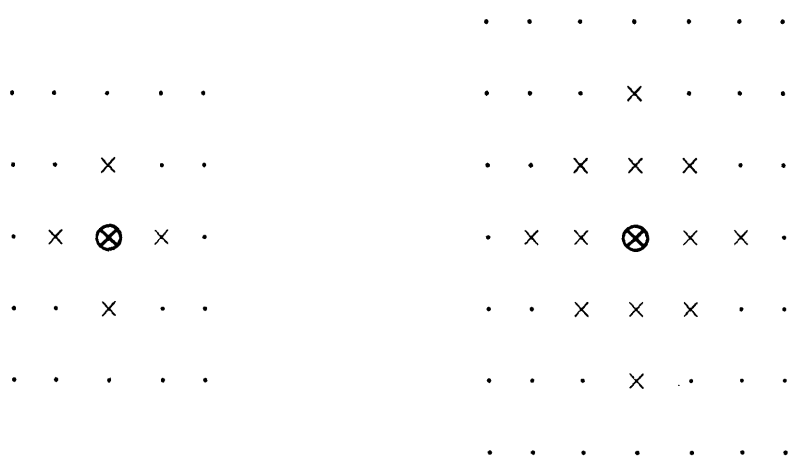
In the information filtering equations (16)–(21), all except (16) preserve sparseness of the information matrix. Also, (16) is the only step that requires an explicit matrix inversion to be performed. Hence, an efficient implementation of the information filter is possible by approximating the prediction step (16) in a way which preserves the sparse matrix structure of the information matrix. Note that while (21) also requires inversion of a matrix, if the information matrices are sparse then this step is just the solution of a sparse system of linear equations, and is hence amenable to both iterative schemes, such as Jacobi and Gauss-Seidel and more sophisticated approaches, such

as multigrid methods [38].

4.1 Spatial Modeling Perspective of the Approximation

As discussed in Section 3.2, an information pair implies a spatial model (23) for the corresponding field estimation error. The information matrix, in particular, can be considered to encode interactions among the components of the field in such a spatial model. Each row of the information matrix $\mathbf{L}(t)$ forms an inner product with the field estimation error vector $\underline{\tilde{f}}(t)$ to yield a weighted average of certain field elements, modeling the interaction among these components of the field.

We intend to constrain the spatial support of such an interaction to be local to a given point as specified by a *neighbor set*, i.e. a connected set of spatial locations within a certain distance (in the Manhattan metric) from the given point in the domain. For example, given a point $\otimes$ on a 2-D lattice, suppose we use $\times$'s to denote the locations of neighbors corresponding to different sets. Then, the nearest neighbor or 1-layer set is shown in the left diagram while the 2-layer set is shown in the right diagram:



(24)

It is physically reasonable that the spatial relationships among the components of the field es-

timation error should be locally defined, as the natural forces and energies governing structural characteristics of the field usually have local extent. Such a constraint of local spatial extent to describe the field interactions corresponds *precisely* to constraining each row of the information matrix to having zeros in all but certain locations, resulting in an overall sparse, diagonally banded structure of the resulting information matrix.

Approximating $\overline{\mathbf{L}}(t)$ by a sparse matrix $\overline{\mathbf{L}}_a(t)$ having a local structure thus corresponds to characterizing the field $\tilde{\underline{\mathbf{f}}}(t)$ by a reduced-order version of the spatial model (23). This insight leads to the following physical intuition for our approximation strategy to the information filter: The prediction step of the information filter can be considered as a model realization process (for the error in the one-step predicted field). Such a model, associated with the *optimal* information filter, tends to yield a full information matrix which characterizes the spatially discrete field by specifying every conceivable interaction among its components. Since the visual field of a natural scene can normally be well specified by spatially local interactions among its components, a reduced order model obtained by approximating the predicted information matrix by one with a given local structure should provide good results.

4.2 Approximating the Information Matrix

Of necessity, the approximated information matrix $\mathbf{L}_a$ corresponding to the reduced order model must have zeros in certain locations and thus must possess a certain, given structure determined solely by the corresponding spatial extent of the allowed field interactions. Once a neighbor structure, such as those in (24), is chosen we need to find a corresponding reduced order information matrix with this structure. The most straightforward approach, and the one we will take, is to simply *mask* the information matrix by zeroing out all elements in prohibited positions. Equiva-

lently we may view this process as the element-by-element multiplication of the information matrix $\mathbf{L}$ by a structuring matrix $\mathcal{W}_\ell$ of zeros and ones so that $\mathbf{L}_a = \mathcal{W}_\ell \odot \mathbf{L}$, where $\odot$ denotes element-by-element multiplication and ℓ is the number of layers in the corresponding neighborhood structure. This masking process corresponds to minimizing the Frobenius norm of the approximation error $\|\mathbf{L} - \mathbf{L}_a\|_F$ over all matrices with the given structure. We term such matrices *$\mathcal{W}_\ell$-structured.*

We could also imagine performing this modeling process through other, more information-theoretic criteria. Recall that an information pair implicitly defines a Gaussian density function, so that we could choose $\mathbf{L}_a$ to be the matrix that minimizes the distance between the Gaussian densities associated with $\mathbf{L}_a$ and $\mathbf{L}$. As our notion of the distance between densities we could use such well known measures as the Bhattacharyya distance or the divergence [35]. Although these approximation criteria are attractive in the sense that they have an information-theoretic foundation, there is no obvious way to compute the structured $\mathbf{L}_a$ easily and efficiently based on them. We thus use the simple truncation approach. Note, however, that we may show that the divergence satisfies the following relationships:

$$\begin{aligned} \text{Divergence}(\mathbf{L}, \mathbf{L}_a) &= \frac{1}{2} \left\| \mathbf{L}_a^{-T/2} (\mathbf{L} - \mathbf{L}_a) \mathbf{L}^{-1/2} \right\|_F^2 \\ &\leq \frac{1}{2} \left(\sum_i \lambda_i^{-1/2}(\mathbf{L}) \right)^2 \left(\sum_i \lambda_i^{-1/2}(\mathbf{L}_a) \right)^2 \|\mathbf{L} - \mathbf{L}_a\|_F^2 \end{aligned} \tag{25}$$

where $\lambda_i(\cdot)$ denote the eigenvalues of the argument. Thus when the information matrix and its approximation are not close to singularity, small values for the Frobenius norm of the matrix approximation error should imply small values of the corresponding divergence. This condition is related to the observability of the field [5], with greater observability being associated with large eigenvalues of the associated information matrices, roughly speaking.

The optimal information filter was given in (16)–(21). A masked information filter can be obtained by replacing (16) with

$$\overline{\mathbf{L}}(t) = \mathcal{W}_\ell \odot \left[\mathbf{U}(t) - \mathbf{U}(t)\mathbf{A}(t)\left(\mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \widehat{\mathbf{L}}(t-1)\right)^{-1}\mathbf{A}^T(t)\mathbf{U}(t)\right] \tag{26}$$

where $\odot$ denotes element-by-element multiplication and $\mathcal{W}_\ell$ is a masking matrix corresponding to the number of layers ℓ in a neighborhood set.

Note that the spatial coherence constraints (7) themselves have a spatial extent associated with them, as reflected in the banded structure of the matrices $\mathbf{S}_i$. If ℓ in (26) is chosen large enough so that the neighborhood set is larger than this spatial coherence extent, then the rest of the filtering algorithm preserves the structure of the information matrix.

4.3 Efficiently Computing the Approximation

The approximation step (26) greatly improves the storage situation, as the number of elements in the approximated information matrix is now only $O(N)$. Further, the inversion operation in the update step (21) can now be implemented efficiently by the afore-mentioned iterative methods (e.g. multigrid methods [38]) due to the sparse structure of the associated matrices. The truncation step (26) by itself, however, does not yet improve on the computational complexity of the optimal information filter because of the inversion of the matrix

$$\mathbf{K}(t) = \mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \widehat{\mathbf{L}}(t-1) \tag{27}$$

on the right hand side of (26). As we have indicated, if $\widehat{\mathbf{L}}(t-1)$ has a banded structure, then so does $\mathbf{K}(t)$. However the same is *not* true of $\mathbf{K}^{-1}(t)$. Thus an algorithm in which we implement (16)

exactly and *then* mask $\overline{\mathbf{L}}(t)$ is computationally prohibitive. Rather what is needed is an efficient method for directly obtaining a banded approximation to $\mathbf{K}^{-1}(t)$ which in turn leads to a banded approximation to $\overline{\mathbf{L}}(t)$ in (16). It is important to emphasize that such an approach involves a second approximation (namely that of $\mathbf{K}^{-1}(t)$) beyond the simple masking of $\overline{\mathbf{L}}(t)$. In the remainder of this section we show how to efficiently propagate such an approximation of the information matrices $\overline{\mathbf{L}}(t)$ and $\hat{\mathbf{L}}(t)$ by efficiently computing a banded approximation to the inverse $\mathbf{K}^{-1}(t)$. The rest of the computations in (26), i.e. matrix multiplications, subtraction, and truncation, are already spatially confined, thus the total computational (and storage) complexity of the resulting algorithm is $O(N)$. Furthermore the sparse banded structure of the calculations allows the possibility of substantial parallelization.

Inversion by Polynomial Approximation

The basis for our approximations of the matrix inverse $\mathbf{K}^{-1}(t)$ of (27) is to express $\mathbf{K}^{-1}(t)$ as an infinite series of easily computable terms. Truncating the infinite series leads us to an efficient computation of the masked information matrix $\overline{\mathbf{L}}(t)$ in (26).

To this end, let us decompose the matrix $\mathbf{K}(t)$, whose inverse we desire, as the sum of two matrices $\mathbf{K} = D + \Omega$, where D is composed of the diagonal of $\mathbf{K}$ and Ω is the remaining, off-diagonal part. Then, if D is invertible, $\mathbf{K}^{-1}$ can be obtained by the following infinite series:

$$\mathbf{K}^{-1} = D^{-1} - D^{-1}\Omega D^{-1} + D^{-1}\Omega D^{-1}\Omega D^{-1} - D^{-1}\Omega D^{-1}\Omega D^{-1}\Omega D^{-1} + \cdots \qquad (28)$$

This series converges if all eigenvalues of $D^{-1}\Omega$ reside within the unit disk, or equivalently, if $\mathbf{K}$ is strictly diagonally dominant [9, 40]. Convergence is especially fast when the eigenvalues of $D^{-1}\Omega$

are close to zero, and taking the first few terms of the series makes a good approximation of the inverse. In fact, if Ω^{-1} exists, it is not difficult to show that

$$\mathbf{K}^{-1} = D^{-1} - D^{-1}\Omega D^{-1} + D^{-1}\Omega\,\mathbf{K}^{-1}\,\Omega D^{-1}. \tag{29}$$

We may use this equation recursively to obtain the series (28). By stopping after a finite number of terms we may thus obtain precise bounds on the error resulting from the use of a truncated series. We may also use the equation to improve the approximation itself by using a coarse approximation to $\mathbf{K}^{-1}$ on the right hand side of (29), for example. For simplicity, however, we will only use here a finite number of terms from (28) for our approximation. Since $\mathbf{K}$ is a matrix with a sparse banded structure in our case, the matrix Ω will also be sparse, leading to a situation where the first several terms in the series are very sparse. The operations involved in computing the finite series approximation are also all locally confined so that they are parallelizable.

The diagonal dominance characteristic of the matrix $\mathbf{K}(t) = \mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \widehat{\mathbf{L}}(t-1)$ is not easy to verify analytically. However, as we mentioned previously, the update step will always tend to increase diagonal dominance of $\mathbf{K}(t)$ thanks to the structure of $\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)$, which itself is generally a *diagonal* matrix of non-negative elements (or a strictly banded matrix with dominant, nonnegative diagonal), thus reinforcing the dominance of the diagonal of $\widehat{\mathbf{L}}$. Since in most computer vision problems $\mathbf{C}(t)$ is calculated on-line, it is necessary to verify the usefulness and accuracy of our approximation through simulations. Later in this and the next section we present several such examples which indicate that our approximation yields excellent, near-optimal results, and we refer the reader to [6] for additional supporting evidence.

4.4 A Sub-Optimal Information Algorithm

We may now present the complete sub-optimal information filter for the system (11), (12) based on the polynomial inverse approximation described in the previous section. The optimal filter is given as (16)–(21). To create a reduced order filter, we first choose a reduced neighborhood interaction structure, thus specifying an associated masking matrix $\mathcal{W}_\ell$. This masking matrix, corresponding to the number of layers in the reduced order model neighborhood, structurally constrains the information matrices. The sub-optimal filter is then obtained by replacing (16) of the optimal filter with the following sequence:

1. Compute the matrix $\mathbf{K}(t) = \mathbf{A}^T(t)\mathbf{U}(t)\mathbf{A}(t) + \widehat{\mathbf{L}}(t-1)$.

2. Decompose $\mathbf{K}(t)$ as $\mathbf{K}(t) = D(t) + \Omega(t)$ where $D(t)$ is composed of the main diagonal of $\mathbf{K}(t)$ and $\Omega(t)$ contains the off diagonal.

3. Use a fixed number of terms in the infinite series (28) to approximate $\mathbf{K}^{-1}(t)$ as $\mathbf{K}_a^{-1}(t)$.

4. $\overline{\mathbf{L}}(t) = \mathcal{W}_\ell \odot \Big[\mathbf{U}(t) - \mathbf{U}(t)\mathbf{A}^T(t)\mathbf{K}_a^{-1}(t)\mathbf{A}(t)\mathbf{U}(t)\Big]$.

4.5 Numerical Results

To examine the effect of our approximations, consider applying the sub-optimal information filter as specified in Section 4.4 to the following dynamic system:

$$\underline{\mathbf{f}}(t) = \underline{\mathbf{f}}(t-1) + \underline{\mathbf{u}}(t), \quad \underline{\mathbf{u}}(t) \sim \langle\langle\, 0, \rho\mathbf{I}\,\rangle\rangle \tag{30}$$

$$\underline{\mathbf{y}}(t) = \begin{bmatrix} \mathbf{I} \\ \mathbf{S}^{(1,0)} \\ \mathbf{S}^{(0,1)} \end{bmatrix} \underline{\mathbf{f}}(t) + \underline{\mathbf{r}}(t), \quad \underline{\mathbf{r}}(t) \sim \left(0, \begin{bmatrix} \nu^{-1}\mathbf{I} & & \\ & \mathbf{I} & \\ & & \mathbf{I} \end{bmatrix} \right) \tag{31}$$

where $\underline{\mathbf{f}}(t)$ is a scalar field defined over a 10×10 spatial domain. Estimation of $\underline{\mathbf{f}}(T)$ corresponds to solving a discrete counterpart of the continuous multi-frame reconstruction problem

$$\min_{\underline{\mathbf{f}}(t)} \int_0^T \int_{\mathcal{D}} \nu \|g - hf\|^2 + \left\| \frac{\partial}{\partial s_1} f \right\|^2 + \left\| \frac{\partial}{\partial s_2} f \right\|^2 + \rho \left\| \frac{\partial}{\partial t} f \right\|^2 \, d\underline{\mathbf{s}} \, dt.$$

Let $\alpha \equiv \rho^{-1}$ and $\beta \equiv \nu^{-1}$. Then, α and β represent the variances of the process and measurement noise processes, respectively. To measure the closeness of approximation of the information matrices we will use the *percent approximation error*, defined as $100 \times \|\mathbf{L}_a - \mathbf{L}_{opt}\| \, / \, \|\mathbf{L}_{opt}\|$, where $\mathbf{L}_a$ is the approximated information matrix and $\mathbf{L}_{opt}$ is the optimal information matrix. The 2-norm [11] is used to compute matrix norms throughout.

Effect of Number of Terms and Structural Constraints

The two charts in Figure 1 show the approximation errors for the predicted and updated information matrices when different numbers of terms are used to approximate the infinite series (28). The filter parameters are $\alpha = \beta = 1$, and the structural constraint is $\mathcal{W}_2$. The six solid lines, from top to bottom, shown in each chart represent the errors when the first one to six terms, respectively, in the series are used. The dashed line in each chart is the error resulting from masking the *exact* matrix inverse (corresponding to an infinite number of series terms) with a $\mathcal{W}_2$ neighborhood structure. As can be observed, as the number of terms increases, the error approaches that corresponding to masking of the exact inverse, although extremely good approximations are obtained with comparatively few terms.

In particular, it appears that the accuracy gained per addition of a term in the series diminishes as the number of terms in the series increases. Here, we quantify such an effect for given structural

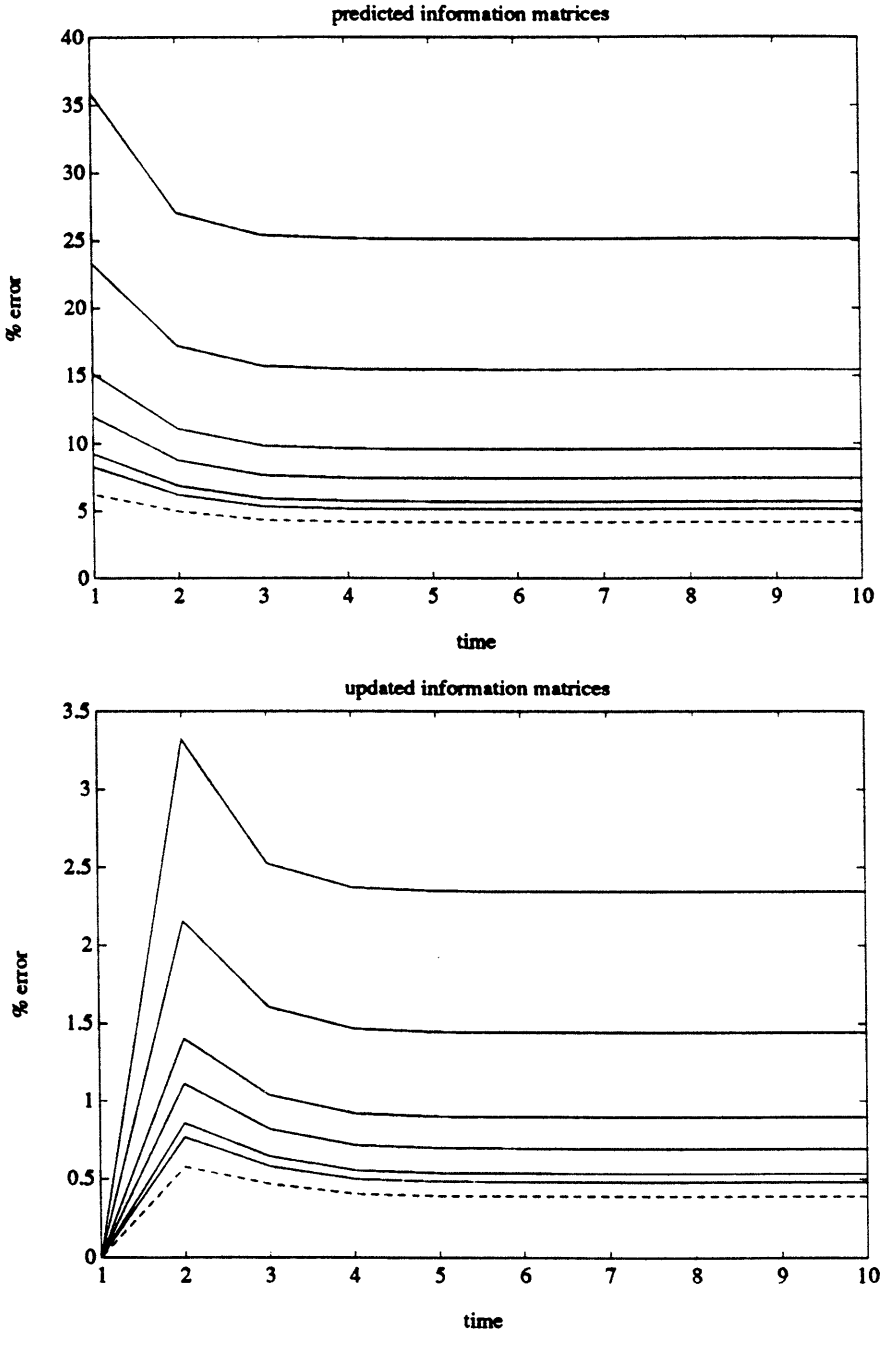


Figure 1: Performance of the sub-optimal information filters using various number of terms in series approximation.

constraints on the information matrices. The two charts in Figure 2 show the approximation errors for the predicted and updated information matrices at $t = 10$ as a function of the number of terms in the series. ($\alpha = \beta = 1.$) The solid lines are the errors associated with a $\mathcal{W}_1$-structural constraint, while the dashed and dotted lines are those associated with $\mathcal{W}_2$ and $\mathcal{W}_3$-structural constraints, respectively. The dash-dot lines represent the errors when no structural constraint is applied. As can be observed, for a tighter structural constraint the gain in accuracy obtained by including more terms in the series levels off at an earlier point.

Effect of Filter Parameters

The effects of the process and measurement noise parameters α and β on the sub-optimal information filter are now determined. Figures 3 and 4 show the errors at $t = 10$ when the number of terms in the series is 2 (solid lines), 4 (dash lines), and 6 (dotted lines). The structural constraint for the information matrices is $\mathcal{W}_2$. The error curves as a function of α (Fig.'s 3 and 4) show unimodal patterns, and the error curves are monotonically increasing with β (Fig. 5).

Summary

A relatively small number of terms in the series (28) is sufficient for an effective approximation of the masked information filter. In particular, a tighter structural constraint $\mathcal{W}_\ell$ on the information matrix allows satisfactory approximation by a smaller number of series terms.

The qualitative effects of the model parameters α and β on the series approximated filter can be explained by the effect of the strength of the process and measurement noises on the structure of the optimal predicted information matrix. When the process noise is progressively decreased, the prediction based on (30) becomes closer to being perfect, and, in particular, the predicted

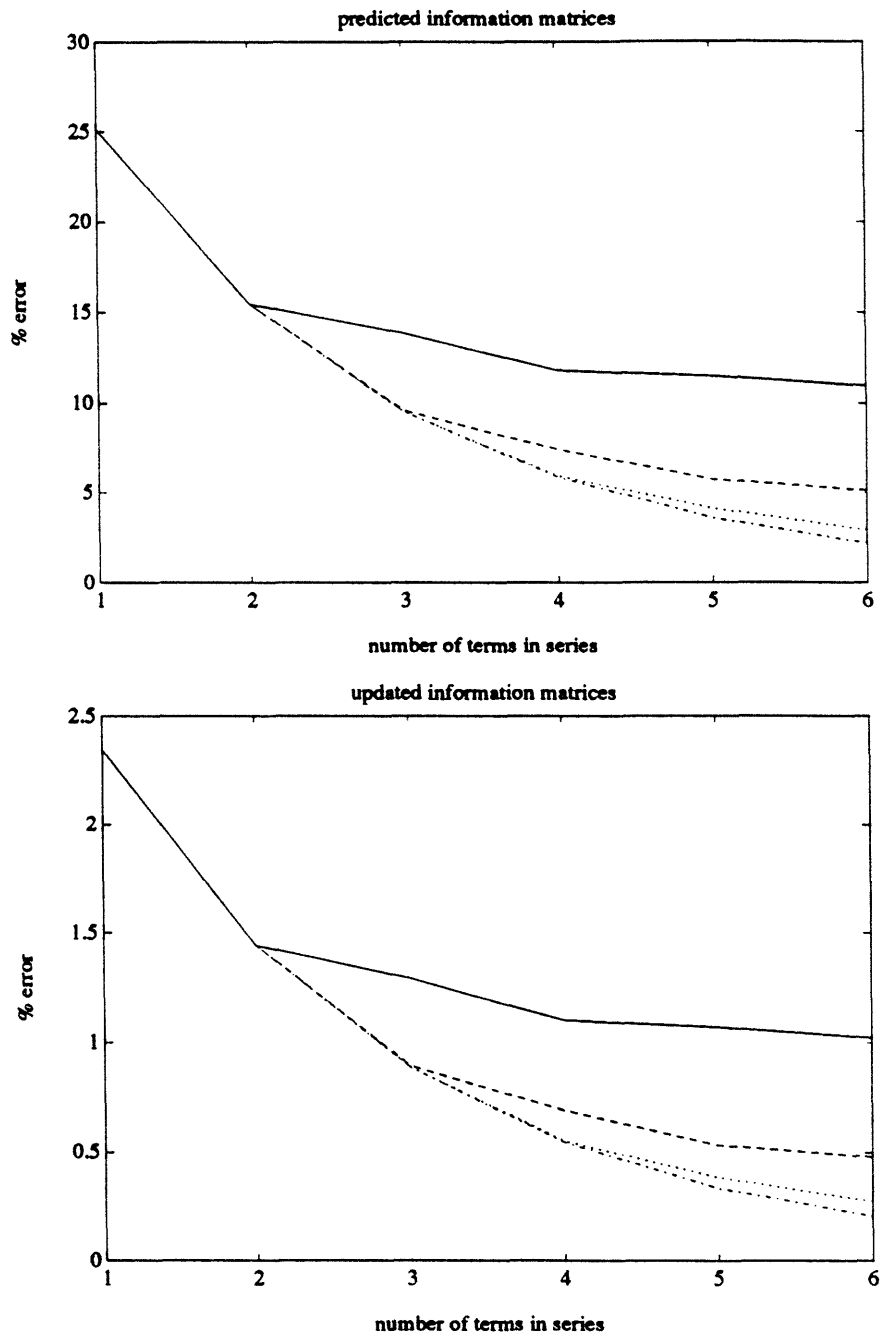


Figure 2: The performance of the series-approximated sub-optimal information filters as a function of the number of terms in the series. The solid, dashed, and dotted lines correspond to the filters with $\mathcal{W}_1$, $\mathcal{W}_2$, and $\mathcal{W}_3$ structural constraints, respectively. The dash-dot lines correspond to the filter with no structural constraint.

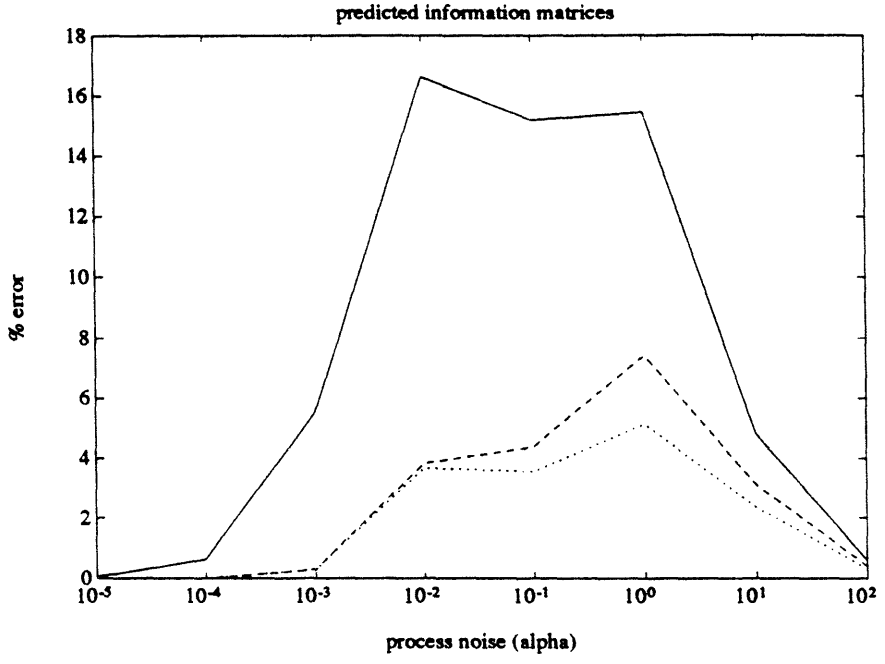


Figure 3: The effect of the process noise parameter on the approximation errors for the predicted information matrices using different numbers of terms in the series approximation — 2 (solid-line), 4 (dashed-line), and 6 (dotted-line).

information matrix approaches the updated information matrix from the previous time frame. Thus, the optimal predicted information matrix in this case almost has the same structure as the updated information matrix, i.e. the nearest neighbor structure and masking has only a small effect. When the process noise is very high, on the other hand, the prediction is close to providing *no* information about the unknown and the optimal predicted information matrix approaches zero. Thus, the structural constraints on the predicted information matrix again has small effect.

The performance of the truncated filters is affected strongly by the strength of the measurement noise. The approximation errors for the predicted information matrices are significantly larger when the measurement noise covariance β is high. Recall that the diagonal information matrix $\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)$ associated with the measurement equation strengthens the diagonal part of the filter information matrix, thereby increasing the relative size of the norm of the elements within the

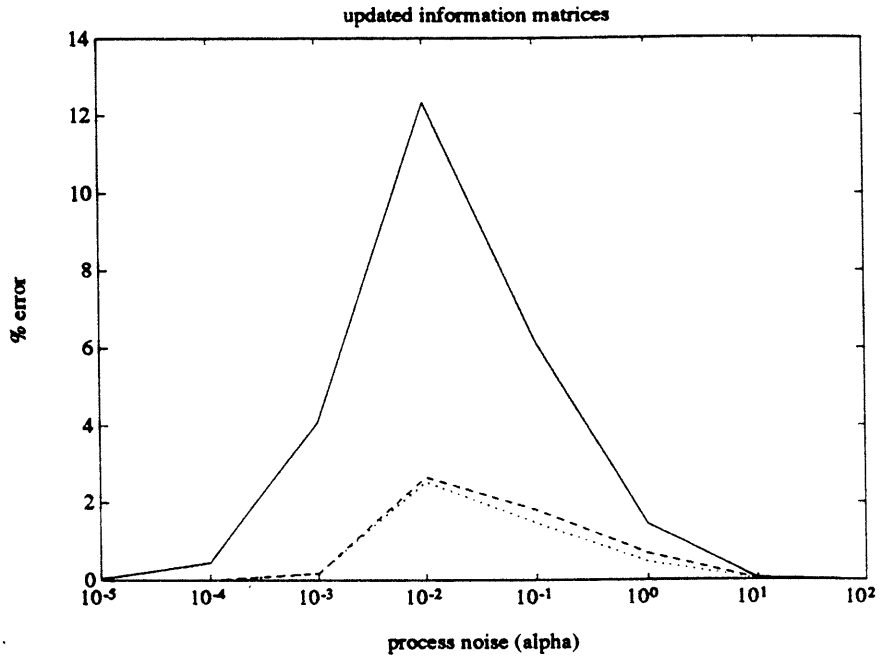


Figure 4: The effect of the process noise parameter on the approximation errors for the updated information matrices using different numbers of terms in the series approximation — 2 (solid-line), 4 (dashed-line), and 6 (dotted-line).

$\mathcal{W}$-structure against the norm of the elements to be truncated. A small value of ν (corresponding to a high level of measurement noise, β), therefore, makes the effect of truncation on the matrix greater. Thus the measurement $\underline{\mathbf{g}}(t)$ of the unknown field $\underline{\mathbf{f}}(t)$ must be modeled to sufficiently high fidelity for the approximation techniques to work.

In this section we have shown that the series-approximated sub-optimal information filter can be used to efficiently approximate the optimal *information matrix*. In the next section we present numerical results on how well this filter produces *estimates* of a visual field $\underline{\mathbf{f}}(t)$.

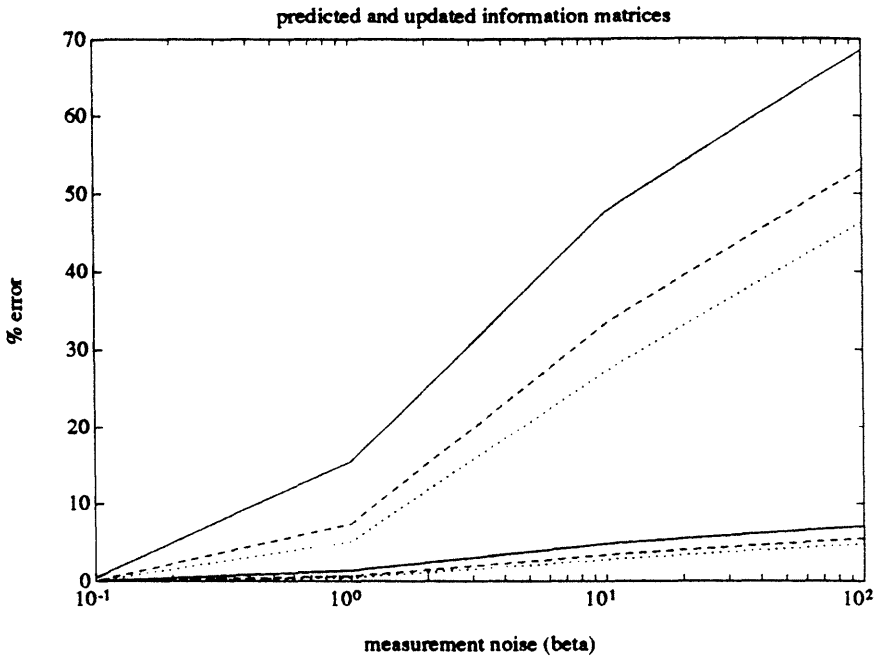


Figure 5: The effect of the measurement noise parameter on the approximation errors for the predicted and updated information matrices using different numbers of terms in the series approximation — 2 (solid-line), 4 (dashed-line), and 6 (dotted-line). The top three lines are associated with the predicted information matrices, while the bottom three lines with the updated information matrices.

5 Simulations: Moving Surface Interpolation

In this section we examine how closely the series approximated information filter of Section 4 can *estimate* artificially generated scalar fields $\underline{\mathbf{f}}(t)$ since this is the final goal of any estimation technique. We add white Gaussian random noise to $\underline{\mathbf{f}}(t)$ to simulate noisy observations $\underline{\mathbf{g}}(t)$ which enter the sub-optimal filters as the inputs. We measure the performance of the sub-optimal Kalman filter through their *% estimation error*

$$\frac{\left\| \mathcal{E}(\,\hat{\underline{\mathbf{f}}}(t)\,) - \underline{\mathbf{f}}(t) \right\|}{\|\underline{\mathbf{f}}(t)\|} \times 100, \tag{32}$$

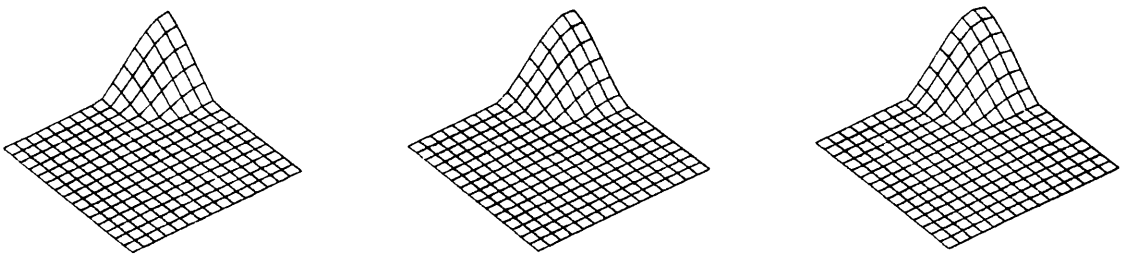


Figure 6: The moving surface to be reconstructed at $t = 2, 4$, and 6.

where $\underline{\hat{f}}(t)$ is the estimate generated by the filters. Each of the sub-optimal filters performs estimation on the same sample path of the observation process $\underline{g}(t)$. The estimates based on several such samples are averaged to obtain an estimate of $\mathcal{E}(\,\underline{\hat{f}}(t)\,)$ for each filter. Our primary concern in this section is to examine how closely the sub-optimal filter can approximate the optimal estimates by comparing the estimation errors (32) associated with the sub-optimal and optimal filters.

5.1 Moving Surface Estimation

A sequence of 16×16 images of the moving tip of a quadratic cone was synthesized and the moving surface reconstructed based on noisy observation of the image sequence using an optimal Kalman filter and series approximated information filter. The actual surface $\underline{f}(t)$ translates across the image frame with a constant velocity whose components along the two frame axes are both 0.2 pixels/frame. That is,

$$f(s_1, s_2, t) = f(s_1 + 0.2, s_2 + 0.2, t - 1).$$

Figure 6 shows $\underline{f}(t)$ at $t = 2, 4$, and 6. Since the spatial coordinates s_1 and s_2 take only integer values in the discrete dynamic model on which the filters are based, we use the following approximate

model

$$\begin{aligned} f(s_1, s_2, t) &= (1-0.2)^2 f(s_1, s_2, t-1) \\ &\quad + (0.2)(1-0.2) f(s_1+1, s_2, t-1) \\ &\quad + (0.2)(1-0.2) f(s_1, s_2+1, t-1) \\ &\quad + (0.2)^2 f(s_1+1, s_2+1, t-1), \end{aligned}$$

which we express as the matrix dynamic equation

$$\underline{\mathbf{f}}(t) = \mathbf{A}\underline{\mathbf{f}}(t-1).$$

In essence, the matrix $\mathbf{A}$ performs approximate spatial shifting of the elements of $\underline{\mathbf{f}}(t-1)$ by a subpixel amount, in this case 0.2 pixels (see, for example, [18] for more details).

A zero-mean white Gaussian process was added to $\underline{\mathbf{f}}(t)$ to simulate a noisy measurement $\underline{\mathbf{g}}(t)$ with SNR of about 2. Moreover, at each t only half of the points of the surface, chosen randomly, were observed. That is, the measurement model is

$$\underline{\mathbf{g}}(t) = \mathbf{H}(t)\underline{\mathbf{f}}(t) + \underline{\mathbf{r}}_0(t) \tag{33}$$

where each entry of $\mathbf{H}(t)$ has 50-50 chance of being 0 or 1 at each time step. This type of partial observation is common in surface interpolation using depth data obtained from stereo matching [13, 14], since matching can be performed only on selected features in the images.

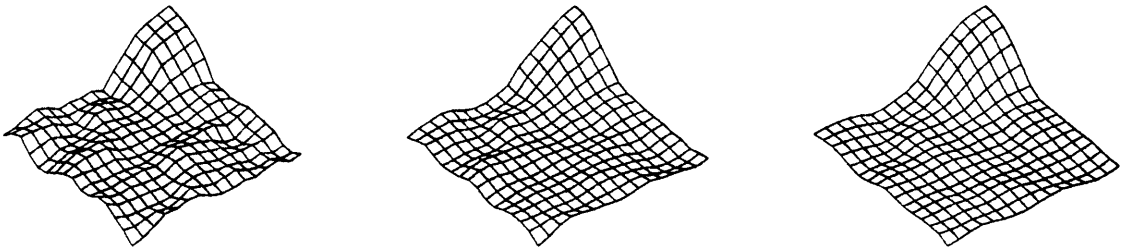


Figure 7: Reconstructed moving surface by optimal Kalman filter at $t = 2, 4$, and 6.

The dynamic system model on which the filters are based is given by:

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{S}^{(1,0)} \\ \mathbf{S}^{(0,1)} \end{bmatrix} \underline{\mathbf{f}}(t) = \begin{bmatrix} \mathbf{I} \\ \mathbf{S}^{(1,0)} \\ \mathbf{S}^{(0,1)} \end{bmatrix} \mathbf{A}\underline{\mathbf{f}}(t-1) + \underline{\mathbf{q}}(t), \quad \underline{\mathbf{q}}(t) \sim (\, 0, \alpha \mathbf{I}\,) \tag{34}$$

$$\begin{bmatrix} \underline{\mathbf{g}}(t) \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{H}(t) \\ \mathbf{S}^{(2,0)} \\ \mathbf{S}^{(0,2)} \\ 2\mathbf{S}^{(1,1)} \end{bmatrix} \underline{\mathbf{f}}(t) + \underline{\mathbf{r}}(t),\ \underline{\mathbf{r}}(t) \sim \left(0, \begin{bmatrix} \beta\mathbf{I} & & & \\ & \mathbf{I} & & \\ & & \mathbf{I} & \\ & & & \mathbf{I} \end{bmatrix} \right). \tag{35}$$

This model corresponds to the use of a thin-plate model for the spatial coherence constraint, as such models are considered particularly suitable for surface interpolation [13]. The dynamic equation reflects the temporal coherence constraint that penalizes large deviation from the dynamic model $\underline{\mathbf{f}}(t) = \mathbf{A}\underline{\mathbf{f}}(t-1)$ *and* imposes smoothness on the deviation $\underline{\mathbf{f}}(t) - \mathbf{A}\underline{\mathbf{f}}(t-1)$ using a membrane model. The application of the membrane model makes the process noise spatially smooth. This assumption is reasonable since the noise reflects (at least partially) the effect of surface motion, which should exhibit some spatial coherence. We let $\alpha = 10^{-2}$ and $\beta = 10^{-1}$.

Figure 7 shows the surfaces reconstructed by the optimal information Kalman filter (16)–(21)

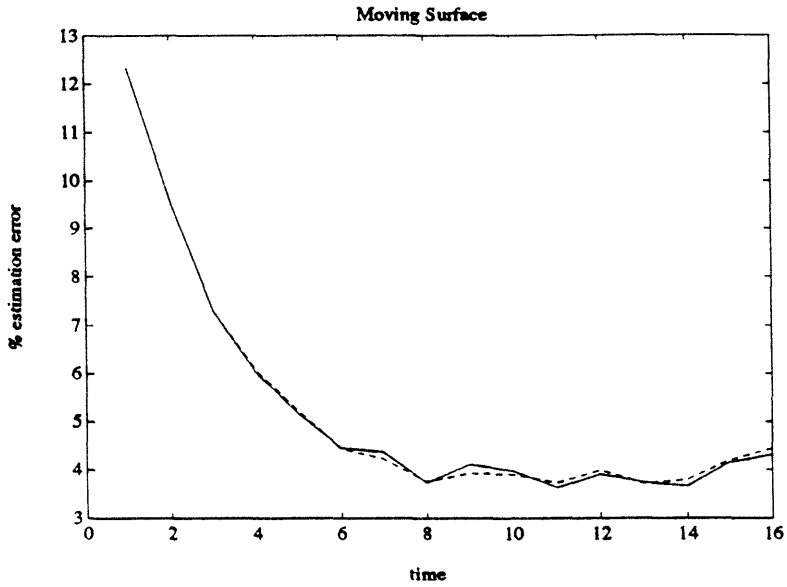


Figure 8: The estimation errors for the optimal Kalman filter (solid line) and the sub-optimal filter (dashed line).

based on the dynamic system above. Observe that the qualitative appearance of the estimated surface improves as more frames of data are incorporated into the estimate. The earlier estimates are expected to be especially noisy because, as indicated by the observation equation, the surface is only partially observable in each image frame.

Figure 8 shows the estimation errors for the optimal Kalman filter (solid line) and series-approximated information filter (dashed line) for the first 16 frames. Four samples paths are averaged to obtain each curve in the figure. The error curves indicate that the sub-optimal filter performs *just as well* as the optimal Kalman filter. The estimation errors for both the optimal and suboptimal filters decrease steadily from about 12% at $t = 1$ to about 4% at $t = 8$. In the series-approximated information filter, the information matrix is constrained to be $\mathcal{W}_6$-structured, and the first 8 terms are used to approximate the infinite series (28) in the prediction step. Such a broader-band approximate model (as compared to a $\mathcal{W}_2$ or nearest-neighbor model) is appropriate

here because of the large spatial extent of the thin-plate model (as opposed to say a membrane model) and the non-zero off-diagonal elements in the system matrix **A**.

5.2 Summary

In this surface reconstruction simulation the approximate filter has performed almost identically to the corresponding optimal Kalman filter. The discrepancy between the optimal filter and the approximate filter appears smaller when the computed estimates (32) are used as a criterion than when the error in the information matrices alone is used, as in the examples in Section 4.5. This property is desirable, since it is the quality of the *estimate* that is of primary concern in the design of approximate filters.

6 Conclusions

We have presented an extension of the classical single-frame visual reconstruction problem by considering the fusing of multiple frames of measurements yielding temporal coherence constraints. The resulting formulation of the multi-frame reconstruction problem is a state estimation problem for the descriptor dynamic system (11) and (12) for which we derived an information filtering algorithm in Section 3.2. Practical limitations arising from the large size of the optimal information matrices led to the development of a sub-optimal scheme. This sub-optimal filter was developed by approximating the field model implied by the optimal information matrix at each step with a reduced order model of fixed spatial extent. This reduced order field model induces a simple structure on the associated information matrices, causing them to be banded and sparse. This structure may be viewed as arising from the imposition of a Markov Random Field structure on the associated visual process. Numerical experiments showed that the resulting sub-optimal filters

provided good approximations to the optimal information matrices and near-optimal estimation performance. Further work is reported in [8, 6], where we present an alternative, *square root* variant of the optimal recursive filter along with an associated near optimal implementation, and in [7], where we apply our filtering results to the sequential estimation of optical flow vector fields and demonstrate the advantages to be obtained in a visual estimation context through the optimal fusing of multiple frames of measurements.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, N.J., 1979.

[2] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, 1988.

[3] G. J. Bierman. *Factorization Methods for Discrete Sequential Estimation.* Academic Press, New York, 1977.

[4] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, Cambridge, Massachesetts, 1987.

[5] R. Brockett. Gramians, generalized inverses, and the least-squares approximation of optical flow. *Journal of Visual Commenication and Image Representation*, 1(1):3–11, 1990.

[6] T. M. Chin. *Dynamic Estimation in Computational Vision.* PhD thesis, Massachusetts Institute of Technology, 1991.

[7] T. M. Chin, W. C. Karl, and A. S. Willsky. Sequential optical flow estimation using temporal coherence. To appear, 1991.

[8] T. M. Chin, W. C. Karl, and A. S. Willsky. A square-root information filter for sequential visual field estimation. To appear, 1991.

[9] P. Concus, G. H. Golub, and G. Meurant. Block preconditioning for the conjugate gradient method. *SIAM J. Sci. Stat. Comput.*, 6:220–252, 1985.

[10] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, 1984.

[11] G. H. Golub and C. F. van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, 1989.

[12] C. D. Greene and B. C. Levy. Smoother implementations for discrete-time Gaussian reciprocal processes. In *Proceedings of 29th IEEE Conference on Decision and Control*, Dec. 1990. Princeton, NJ.

[13] W. E. L. Grimson. A computational theory of visual surface interpolation. *Proceedings of the Royal Society of London B*, 298:395–427, 1982.

[14] W. E. L. Grimson. An implementation of a computational theory of visual surface interpolation. *Computer Vision, Graphics, and Image Processing*, 22:39–69, 1983.

[15] N. M. Grzywacz, J. A. Smith, and A. L. Yuille. A common theoretical framework for visual motion's spatial and temporal coherence. In *Proceedings of Workshop on Visual Motion*, pages 148–155. IEEE Computer Society Press, 1989. Irvine, CA.

[16] J. Heel. Dynamic motion vision. In *Proceedings of the DARPA Image Understanding Workshop*, 1989. Palo Alto, CA.

[17] J. Heel. Direct estimation of structure and motion from multiple frames. A.I.Memo No. 1190, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.

[18] J. Heel. *Temporal Surface Reconstruction.* PhD thesis, Massachusetts Institute of Technology, 1991.

[19] J. Heel and S. Rao. Temporal integration of visual surface reconstruction. In *Proceedings of the DARPA Image Understanding Workshop*, 1990. Pittsburgh, PA.

[20] F. Heitz, P. Perez, E. Memin, and P. Bouthemy. Parallel visual motion analysis using multiscale Markov random fields. In *Proceedings of Workshop on Visual Motion.* IEEE Computer Society Press, 1991. Princeton, NJ.

[21] E. C. Hildreth. Computations underlying the measurement of visual motion. *Artificial Intelligence*, 23:309–354, 1984.

[22] B. K. P. Horn. Image intensity understanding. *Artificial Intelligence*, 8:201–231, 1977.

[23] B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, Massachesetts, 1986.

[24] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[25] K. Ikeuchi. Determination of surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3:661–669, 1981.

[26] K. Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.

[27] A. H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[28] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

[29] B. C. Levy, R. Frezza, and A. J. Krener. Modeling and estimation of discrete-time Gaussian reciprocal processes. to appear in IEEE Transactions on Automatic Control, 1990.

[30] F. L. Lewis. *Optimal Estimation.* John Wiley & Sons, New York, 1986.

[31] L. H. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3, 1989.

[32] R. Nikoukhah. *A Deterministic and Stochastic Theory for Two-point Boundary-value Descriptor Systems.* PhD thesis, Massachusetts Institute of Technology, 1988.

[33] R. Nikoukhah, A. S. Willsky, and B. C. Levy. Kalman filtering and Riccati equations for descriptor systems. submitted to IEEE Transactions on Automatic Control, 1991.

[34] A. Rougee, B. C. Levy, and A. S. Willsky. An estimation-based approach to the reconstruction of optical flow. Technical Report LIDS-P-1663, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1987.

[35] F. C. Schweppe. *Uncertain Dynamic Systems.* Prentice-Hall, Englewood Cliffs, N.J., 1973.

[36] A. Singh. Incremental estimation of image-flow using a Kalman filter. In *Proceedings of Workshop on Visual Motion*, pages 36–43. IEEE Computer Society Press, 1991. Princeton, NJ.

[37] R. Szeliski. *Baysian Modeling of Uncertainty in Low-level Vision.* Kluwer Academic Publishers, Norwell, Massachuesetts, 1989.

[38] D. Terzopoulos. Image analysis using multigrid relaxation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:129–139, 1986.

[39] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:413–424, 1986.

[40] R. S. Varga. *Matrix Iterative Analysis.* Prentice-Hall, Englewood Cliffs, N.J., 1962.

[41] J. W. Woods. Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theorey*, IT-18:232–240, 1972.

Contents