# OPTIMAL BRIGHTNESS FUNCTIONS FOR OPTICAL FLOW ESTIMATION OF DEFORMABLE MOTION 

August 26, 1997

Thomas S. Denney Jr. and Jerry L. Prince

## Affiliations:

Thomas S. Denney Jr.:
Image Analysis and Communications Laboratory
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218
Jerry L. Prince:
Image Analysis and Communications Laboratory
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218
Correspond to:
Jerry L. Prince
105 Barton Hall
The Johns Hopkins University
Baltimore, MD 21218
Tel: (410) 516-5192
Fax: (410) 516-5566
E-mail: prince@mashie.ece.jhu.edu

## Acknowledgment of Support:

This research was supported by the Whitaker Foundation of Mechanicsburg, PA.

IP Editors' Information Classification Scheme (EDICS):
1.11

# OPTIMAL BRIGHTNESS FUNCTIONS FOR OPTICAL FLOW ESTIMATION OF DEFORMABLE MOTION 

Thomas S. Denney Jr. and Jerry L. Prince


#### Abstract

Estimation accuracy of Horn and Schunck's classical optical flow algorithm depends on many factors including the brightness pattern of the measured images. Since some applications can select brightness functions with which to "paint" the object, it is desirable to know what patterns will lead to the best motion estimates. In this paper we present a method for determining this pattern a priori using mild assumptions about the velocity field and imaging process. Our method is based on formulating Horn and Schunck's algorithm as a linear smoother and rigorously deriving an expression for the corresponding error covariance function. We then specify a scalar performance measure and develop an approach to select an optimal brightness function which minimizes this performance measure from within a parametrized class. Conditions for existence of an optimal brightness function are also given. The resulting optimal performance is demonstrated using simulations, and a discussion of these results and potential future research is given. T.S. Denney Jr. is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218. J.L. Prince is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218.

Permission to publish this abstract separately is granted.


## I. INTRODUCTION

The algorithm developed by Horn and Schunck [1] for estimating the optical flow between image pairs, which we will refer to as standard optical flow (SOF), has been widely studied in the computer vision community. It is generally accepted that SOF produces a good overall qualitative picture of the motion field, but lacks good quantitative behavior, especially when the images involve rigid body motion with possible occlusion [2, 3]. In these cases, parametric methods such as those reported in $[4,5,6,7,8,9,10]$ and modified Horn and Schunck methods such as those reported in $[2,3,11,12,13,14,15,16,10]$ show superior quantitative performance. When the images show an object undergoing deformable motion with no occlusion, however, SOF may still provide a high-resolution, accurate estimate of the motion field. In these cases there are many parameters affecting the performance of SOF including spatial and temporal sampling, the regularization coefficient, the nature of the motion and - what is of primary interest in this paper - the spatial pattern of brightness of the object itself.

In general, one cannot control the spatial pattern of brightness of the object within an image sequence since it is an inherent part of the underlying physics and imaging process. In some applications, however, it is possible to control this brightness function. For example, consider the estimation of left ventricular motion from a sequence of magnetic resonance (MR) images of the heart. Recent developments in MR tagging [17, 18, 19, 20] make it possible to modulate the MRI brightness function to make a spatial pattern appear in otherwise homogeneous tissue (see Section VI for more information). Prince et al. [21, 22, 23, 24] have shown that such patterns can be exploited using optical flow processing to detect motion that would otherwise be obscured by the aperture problem (cf. [25]). This work also revealed that the performance of SOF is strongly affected by the spatial frequency of the spatial pattern placed in the images. This observation leads naturally to the general question: what brightness function results in the best estimate of motion given that SOF is used to process the image pairs? In this paper we deal with the somewhat more restricted problem of the a priori selection of the parameters that will optimize SOF performance given a parameterized class of brightness functions. We call the pattern specified by the optimal parameters the optimal brightness function.

The primary difficulty in determining the optimal brightness function is the development of a measure of SOF performance. Horn and Schunck's optical flow algorithm is based on a variational formulation that has no inherent performance measure. Several error analyses of optical flow and related motion estimation procedures have been formulated for rigid-body
motion [26, 4, 2, 5]; however, these results do not apply to deformable objects. Kearney, et al. [27, 28] discussed error sources in SOF which depend on the brightness function and proposed heuristic methods for quantifying these errors. These results, however, do not provide a rigorous framework for the development of an overall performance measure. Simoncelli, et. al. [29] developed a gradient-based estimation algorithm for general motion which uses Gaussian models for the velocity field and noise sources, and this algorithm provides a performance measure for the resulting velocity estimate. Their algorithm, however, is not SOF. Chin [30] derived a discrete version of Horn and Schunck's optical flow equations for deterministic motions in 2D using a maximum-likelihood (ML) approach, and this approach does provide an expression for the estimation error covariance which can be used as a performance measure. As part of our development in this paper, we present an alternate derivation of Chin's error covariance starting with the linear smoothing formulation of Rougee et al. [31, 32]. We show in Section IV, however, that for the estimation error covariance to be an accurate performance measure, a new measurement noise model must be developed.

In this paper, we use the estimation error covariance and a new measurement noise model to develop a criterion for brightness function optimality, and we develop a method to select the optimal brightness function from a parametrized class of functions. Knowledge of (or assumptions about) the velocity field smoothness, maximum velocity, and imaging noise variance are required; however, this information is generally available or easily deduced in any application. Furthermore, we show empirically that the optimal brightness function choice is relatively robust to modeling errors. Our general approach was previously reported in [33], where we considered a one-dimensional analog to the optical flow problem. In this paper, however, we present a comprehensive treatment of brightness function optimization in two dimensions, a subject which to our knowledge has not been previously addressed in the literature.

This paper is organized as follows. In Section II we present some background on the motion of deformable objects and on the linear smoothing formulation of SOF. In Section III we derive an expression for the optical flow error covariance based on this optimal linear smoother. In Section IV we develop an a priori performance measure for SOF based on the error covariance and describe how to determine the optimal brightness function. We present some simulation results in Section V and provide a discussion of our results and of possible future research directions in Section VI.

## II. BACKGROUND

## A. Motion Model

Since we are concerned with estimating the motion of deformable objects, we will use notation and terminology from continuum mechanics [34]. In this theory, the body is the object undergoing motion; it consists of material points, which may be thought of as small physical particles. A motion as shown in Figure 1 is a function that maps the material points to spatial points in the image at time $t$. We write $\mathbf{r}=\mathbf{r}(\mathbf{p}, t)$ to indicate that the material point $\mathbf{p}$ has moved to the spatial point $\mathbf{r}$ at time $t$. There exists an inverse function called the reference map (see Figure 1) which gives the material point for each spatial point in the image at time $t$. We write $\mathbf{p}=\mathrm{p}(\mathbf{r}, t)$ to indicate that the spatial point $\mathbf{r}$ at time $t$ corresponds to the material point $\mathbf{p}$. The reference map and motion map satisfy

$$
\begin{align*}
\mathrm{r}(\mathrm{p}(\mathbf{r}, t), t) & =\mathbf{r}  \tag{1a}\\
\mathrm{p}(\mathrm{r}(\mathbf{p}, t), t) & =\mathbf{p} \tag{1b}
\end{align*}
$$

The spatial velocity $v(\mathbf{r}, t)=[\mu(\mathbf{r}, t), \nu(\mathbf{r}, t)]^{T}$ is the function estimated by SOF; it is related to the motion map by the equation

$$
\begin{equation*}
\frac{\partial}{\partial t} \mathrm{r}(\mathbf{p}, t)=v(\mathrm{r}(\mathbf{p}, t), t) \tag{2}
\end{equation*}
$$

The image of the body at time $t$ is described by a brightness function $\varphi(\mathbf{r}, t)$. At time $t=0$, the body is said to be in the reference configuration, which means that the spatial points and the material points are identical, i.e.

$$
\begin{aligned}
& \mathrm{r}(\mathbf{p}, 0)=\mathbf{p} \\
& \mathrm{p}(\mathbf{r}, 0)=\mathbf{r} .
\end{aligned}
$$

Therefore, denoting the image of the body in the reference configuration as $f(\mathbf{p})$, it follows that

$$
f(\mathbf{p})=\left.\varphi(\mathbf{r}, 0)\right|_{\mathbf{r}=\mathbf{p}}
$$

and

$$
\varphi(\mathbf{r}, t)=f(\mathrm{p}(\mathbf{r}, t))
$$

As an example, consider the reference configuration

$$
f(\mathbf{p})=\sin \left(\omega p_{x}\right) \sin \left(\omega p_{y}\right)
$$

and the constant velocity $\mathbf{v}=[\mu \nu]^{T}$. It follows that the motion map is

$$
\mathbf{r}=\mathbf{r}(\mathbf{p}, t)=\mathbf{p}+\mathbf{v} t
$$

the reference map is

$$
\mathbf{p}=\mathrm{p}(\mathbf{r}, t)=\mathbf{r}-\mathbf{v} t
$$

and the brightness function is

$$
\varphi(\mathbf{r}, t)=f(\mathrm{p}(\mathbf{r}, t))=\sin \left(\omega\left(r_{x}-\mu t\right)\right) \sin \left(\omega\left(r_{y}-\nu t\right)\right)
$$

Finally, we note that

$$
\frac{\partial}{\partial t} \mathrm{r}(\mathrm{p}(\mathbf{r}, t))=\mathbf{v}
$$

which agrees with Equation (2).

## B. Variational Formulation

In general, the motion estimation problem involves estimating the velocity $v(\mathbf{r}, t)$ of each point in an image from the brightness function $\varphi(\mathbf{r}, t)$. If the brightness of each material point $\mathbf{p}$ is constant in time, using the chain rule to differentiate $\varphi(\mathbf{r}, t)$ with respect to time while holding the material coordinate $\mathbf{p}$ constant [34] yields the Brightness Constraint Equation (BCE) (cf. [1])

$$
\begin{equation*}
\nabla \varphi(\mathbf{r}, t) \cdot v(\mathbf{r}, t)+\varphi_{t}(\mathbf{r}, t)=0 \tag{4}
\end{equation*}
$$

where the subscript $t$ denotes partial differentiation with respect to $t$. Note that in the continuous case, the BCE is exact [34]. We address the effects of discretization in Section IV. In two dimensions, solving (4) for the velocity is an ill-posed problem because there are two components to the velocity at each spatial point and only one (linear) equation relating these components. Horn and Schunck [1] solved this problem by imposing a spatial regularity condition and using the calculus of variations to show that the two components of the velocity estimate satisfy

$$
\begin{align*}
\nabla^{2} \hat{\mu} & =\frac{1}{\alpha^{2}} \varphi_{x}\left[\varphi_{x} \hat{\mu}+\varphi_{y} \hat{\nu}+\varphi_{t}\right]  \tag{5a}\\
\nabla^{2} \hat{\nu} & =\frac{1}{\alpha^{2}} \varphi_{y}\left[\varphi_{x} \hat{\mu}+\varphi_{y} \hat{\nu}+\varphi_{t}\right] \tag{5b}
\end{align*}
$$

where $\varphi_{x}$ and $\varphi_{y}$ denote partial differentiation of $\varphi$ with respect to $x$ and $y$ respectively. Typically, these coupled Poisson equations are discretized, and the resulting large linear system of equations represented by

$$
\begin{equation*}
\Sigma \hat{V}=Y \tag{6}
\end{equation*}
$$

is solved by simultaneous over relaxation (SOR), Gauss-Seidel, or related relaxation methods (cf. [35]). While the variational formulation of the SOF algorithm is relatively straightforward, it does not provide a way to predict the quality of the resulting velocity estimate. The stochastic formulation developed below, however, will allow us to show in Section III that the quality of the SOF estimate is ultimately determined by the matrix $\Sigma$.

## C. Stochastic Formulation

Rougee et al. [31, 32] showed that when the motion is modeled as a particular boundary value random process and the multi-dimensional linear smoothing methods of Adams et al. [36] are employed, the optimal linear smoother is also given by (5). This derivation is particularly important in our work because the linear smoother has analytic expressions for the velocity estimation error which we use to derive a performance measure. In this section we give an alternate derivation of Rougee et al.'s result which closely follows the work of Adams et al. [36]. In the process we provide a notation and conceptual framework that is used in subsequent sections.

Assume that the velocity is defined on a regular domain $\Omega$ with boundary $\partial \Omega$ and has a state model of the form

$$
\begin{align*}
L v(x, y, t)=u(x, y, t) & \text { for }(x, y) \in \Omega  \tag{7a}\\
F v_{b}(s)=u_{b}(s) & \text { for } s \in \partial \Omega \tag{7b}
\end{align*}
$$

where $u \sim N\left(0, \sigma_{u}^{2} I\right), u_{b}(s) \sim N\left(0, \sigma_{u}^{2}\right), L=I \otimes \nabla$, and $F=d / d s$. Here, the symbol $\otimes$ means Kronecker product and the subscript $b$ implies a restriction of the variable to the boundary $\partial \Omega$. We now drop the explicit notation indicating spatial and temporal dependence. The observations are a noisy version of the BCE [Equation (4)]

$$
\begin{align*}
y=C v+w & \text { on } \Omega  \tag{8a}\\
y_{b}=C v_{b}+w_{b} & \text { on } \partial \Omega \tag{8b}
\end{align*}
$$

where $w \sim N\left(0, \sigma_{w}^{2}\right)$ and $w_{b} \sim N\left(0, \sigma_{w}^{2}\right)$. To match the BCE, the output gain is the spatial brightness gradient $C=\nabla \varphi$, and the measurements are the negative temporal brightness gradient $y=-\varphi_{t}$ and $y_{b}=-\varphi_{t}$. The adjoint of $L$, denoted by $L^{\dagger}$, is defined by Green's Identity [36]

$$
\begin{equation*}
\langle L x, \lambda\rangle_{L_{2}^{2}(\Omega)}=\left\langle x, L^{\dagger} \lambda\right\rangle_{L_{2}^{2}(\Omega)}+\left\langle x_{b}, E \lambda_{b}\right\rangle_{H_{b}}, \tag{9}
\end{equation*}
$$

where $L_{2}^{2}(\Omega)$ and $H_{b}$ are Hilbert spaces of 2-vector square-integrable functions defined on $\Omega$ and $\partial \Omega$ respectively. It can be shown that the adjoint of $L$ is given by the modified
divergence operator $L^{\dagger}=-I \otimes \nabla \cdot$ and that $E=I \otimes n^{T}(s)$, where $n(s)$ is the unit vector normal to $\partial \Omega$ at the boundary point $s$.

Identifying $\lambda$ as the complementary process of $v$ (cf. [36]), and $\hat{\lambda}$ and $\hat{v}$ as their respective estimates, the estimate Hamiltonian is given by

$$
\begin{align*}
& {\left[\begin{array}{cc}
L & -\sigma_{u}^{2} I \\
\frac{1}{\sigma_{w}^{2}} C^{T} C & L^{\dagger}
\end{array}\right]\left[\begin{array}{l}
\hat{v} \\
\hat{\lambda}
\end{array}\right]=\left[\begin{array}{c}
0 \\
\frac{1}{\sigma_{w}^{2}} C^{T} y
\end{array}\right] \quad \text { on } \Omega}  \tag{10a}\\
& {\left[\begin{array}{cc}
\frac{1}{\sigma_{w}^{2}} C^{T} C+\frac{1}{\sigma_{u}^{2}} F^{*} F & E
\end{array}\right]\left[\begin{array}{c}
\hat{v}_{b} \\
\hat{\lambda_{b}}
\end{array}\right]=\frac{1}{\sigma_{w}^{2}} C^{T} y_{b} \quad \text { on } \partial \Omega \text {. }} \tag{10b}
\end{align*}
$$

The estimate equation can be simplified by substituting the top row of (10a)

$$
\hat{\lambda}=\frac{1}{\sigma_{u}^{2}} L \hat{v}
$$

into the bottom row of (10a) and into (10b) (with a restriction to the boundary) yielding

$$
\begin{align*}
& {\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C+L^{\dagger} L\right] \hat{v}=\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} y \quad \text { on } \Omega}  \tag{11a}\\
& {\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C+F^{*} F+E L\right] \hat{v}_{b}=\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} y_{b} \quad \text { on } \partial \Omega .} \tag{11b}
\end{align*}
$$

Note that $E L \hat{v}_{b}=I \otimes n^{T} \nabla \hat{v}_{b}$, the normal derivative of $\hat{v}_{b}$. Since $L^{\dagger} L=-I \otimes \nabla^{2}$, (11a) is exactly (5), the optical flow equation of Horn and Schunck [1] with $\alpha^{2}=\sigma_{w}^{2} / \sigma_{u}^{2}$. Thus the variational formulation and the stochastic formulation are equivalent in the sense that the same set of equations must be solved to yield the velocity estimate. In addition, because (11) results in the minimum mean-square-error (MSE) estimate, $\alpha^{2}$ can be interpreted as the optimal regularization parameter. Finally, note that if the boundary smoothness is unknown and there are no measurements on the boundary, the boundary conditions in (11b) reduce to the standard Neumann boundary conditions used by Horn and Schunck [1].

## III. Error Covariance

In this section we use the stochastic formulation of Section II to derive an expression for the optical flow performance. We first derive equations for the continuous estimation error covariance using the complementary model methods of Adams, et al. [36]. We then discretize the error equations and put them in the nearest neighbor model (NNM) form of [37]. Next we develop an expression for the discrete error covariance. Finally, we specify a scalar performance measure based on the discrete error covariance and describe its calculation.

## A. Estimation Error

We define the velocity estimation error as $\tilde{v}=\hat{v}-v$. The Hamiltonians involving the velocity error on the interior and boundary are given by (cf. [36])

$$
\left.\begin{array}{r}
{\left[\begin{array}{cc}
L & -\sigma_{u}^{2} I \\
\frac{1}{\sigma_{w}^{2}} C^{T} C & L^{\dagger}
\end{array}\right]\left[\begin{array}{c}
\tilde{v} \\
-\hat{\lambda}
\end{array}\right]=\left[\begin{array}{c}
u \\
-\frac{1}{\sigma_{w}^{2}} C^{T} w
\end{array}\right] \quad \text { on } \Omega} \\
{\left[\frac{1}{\sigma_{w}^{2}} C^{T} C+\frac{1}{\sigma_{u}^{2}} F^{*} F\right.}  \tag{12b}\\
E
\end{array}\right]\left[\begin{array}{c}
\tilde{v}_{b} \\
-\hat{\lambda}_{b}
\end{array}\right]=\frac{1}{\sigma_{u}^{2}} F^{*} u_{b}-\frac{1}{\sigma_{w}^{2}} C^{T} w_{b} \quad \text { on } \partial \Omega . .
$$

Solving the top row of (12a) for $\tilde{\lambda}$ and substituting into the bottom row of (12a) and into (12b) yields

$$
\begin{align*}
& {\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C+L^{\dagger} L\right] \tilde{v}=L^{\dagger} u-\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} w \quad \text { on } \Omega}  \tag{13a}\\
& {\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C+F^{*} F+E L\right] \tilde{v}_{b}=\left[F^{*}+E\right] u_{b}-\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} w_{b} . \quad \text { on } \partial \Omega .} \tag{13b}
\end{align*}
$$

It is possible to derive an expression for the continuous error covariance based on (13), but this approach presents problems when the continuous error covariance is discretized on a square lattice. Instead, our approach is to derive an expression for the discrete error covariance directly from a discretized version of (13).

## B. Discretization

We assume the velocity is defined on an $N \times N$ lattice $\Omega$ as shown in Figure 2. The boundary $\Omega_{b}$ of this lattice is an index set containing the first and last two rows and columns of $\Omega$, and the interior $\tilde{\Omega}$ is an index set of the remaining points of $\Omega$.

The general form of a nearest neighbor discrete operator $L$ is [37]

$$
\begin{equation*}
L x_{i j}=\left[A_{0}+A_{1} D_{1}+A_{2} D_{1}^{-1}+A_{3} D_{2}+A_{4} D_{2}^{-1}\right] x_{i j} \tag{14}
\end{equation*}
$$

where $D_{1} x_{i j}=x_{i-1, j}$ and $D_{2} x_{i j}=x_{i, j-1}$. The adjoint of $L, L^{\dagger}$, is defined using the discrete form of Green's Identity [37]

$$
\begin{equation*}
\langle L x, \lambda\rangle_{S(\tilde{\Omega})}=\left\langle x, L^{\dagger} \lambda\right\rangle_{S(\tilde{\Omega})}+\left\langle x_{b}, E \lambda_{b}\right\rangle_{S_{b}} \tag{15}
\end{equation*}
$$

where $S(\tilde{\Omega})$ and $S_{b}$ are the vector spaces of 2-vector functions defined over the index sets $\tilde{\Omega}$ and $\Omega_{b}$ respectively. The matrix $E$ is determined by the domain $\Omega$ and the operator $L$ [36] and is explicitly defined in Appendix A. The resulting $L^{\dagger}$ is defined by

$$
\begin{equation*}
L^{\dagger} x_{i j}=\left[A_{0}^{T}+A_{2}^{T} D_{1}+A_{1}^{T} D_{1}^{-1}+A_{4}^{T} D_{2}+A_{3}^{T} D_{2}^{-1}\right] x_{i j} \tag{16}
\end{equation*}
$$

For optical flow, we define $L$ as the forward difference gradient operating on each component of the velocity. Thus,

$$
A_{0}=\left[\begin{array}{l}
-I \\
-I
\end{array}\right], \quad A_{2}=\left[\begin{array}{l}
I \\
0
\end{array}\right], \quad A_{4}=\left[\begin{array}{l}
0 \\
I
\end{array}\right]
$$

and $A_{1}=A_{3}=0$. Here, and in subsequent expressions, the identity matrix $I$ without a subscript denotes the $2 \times 2$ identity matrix. Equation (16) simplifies to

$$
\begin{equation*}
L^{\dagger} x_{i j}=\left[A_{0}^{T}+A_{2}^{T} D_{1}+A_{4}^{T} D_{2}\right] x_{i j} \tag{17}
\end{equation*}
$$

and it follows that

$$
\begin{align*}
L^{\dagger} L x_{i j} & =\left[A_{0}^{T}+A_{2}^{T} D_{1}+A_{4}^{T} D_{2}\right]\left[A_{0}+A_{2} D_{1}^{-1}+A_{4} D_{2}^{-1}\right] \\
& =\left[4 I-I D_{1}-I D_{1}^{-1}-I D_{2}-I D_{2}^{-1}\right] x_{i j} . \tag{18}
\end{align*}
$$

Note that $L^{\dagger} L$ is the negative of the discrete Laplacian mask

$$
\nabla^{2} \approx\left[\begin{array}{ccc} 
& +I & \\
+I & -4 I & +I \\
& +I &
\end{array}\right]
$$

which is often used to solve the coupled Poisson equations (5) which define the solution to SOF. The boundary vector $X_{b}$ is defined as [37]

$$
X_{b}=\left[\begin{array}{c}
X_{1}  \tag{19}\\
X_{2} \\
X_{N} \\
X_{N-1} \\
x_{1}^{\prime} \\
x_{2}^{\prime} \\
x_{N}^{\prime} \\
x_{N-1}^{\prime}
\end{array}\right] \quad \text { where } \quad X_{i}=\left[\begin{array}{c}
x_{i 1} \\
x_{i 2} \\
\vdots \\
x_{i N}
\end{array}\right] \quad \text { and } \quad x_{j}^{\prime}=\left[\begin{array}{c}
x_{2 j} \\
x_{3 j} \\
\vdots \\
x_{N-1 j}
\end{array}\right]
$$

Note that $X_{i}$ is the entire $i$ th row of the image and $x_{j}^{\prime}$ is the $j$ th column of the image with the first and last pixels removed. Also, note that the entries $x_{21}, x_{22}, x_{N-11}, x_{N-12}, x_{2 N}$, $x_{2 N-1}, x_{N-1 N}$ and $x_{N-1 N-1}$ appear twice in $X_{b}$.

## C. Discrete Error Covariance

After some work we find that the discrete versions of the error equations (13) are given by

$$
\begin{gather*}
{\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C+L^{\dagger} L\right] \tilde{v}_{i j}=L^{\dagger} u_{i j}-\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} w_{i j}}  \tag{20a}\\
{\left[\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} H^{T} H+F^{T} F+E L\right] \tilde{V}_{b}=\left(F^{T} \Delta_{F}+E\right) U_{b}-\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} H^{T} W_{b}} \tag{20b}
\end{gather*} \quad \text { otherwise }
$$

where $H=I_{2(8 N-8)} \otimes C$ and the matrices $F$ and $\Delta_{F}$ are defined in Appendix A. Equations (20a) and (20b) can also be expressed as

$$
\begin{equation*}
\Sigma \tilde{V}=\Sigma_{u} U-\Sigma_{w} W \tag{21}
\end{equation*}
$$

where

$$
\tilde{V}=\left[\begin{array}{c}
\tilde{V}_{1} \\
\tilde{V}_{2} \\
\vdots \\
\tilde{V}_{N}
\end{array}\right]
$$

and $U$ and $W$ are similarly defined. Note that $\Sigma$ is the same matrix used in the estimate equation

$$
\begin{equation*}
\Sigma \hat{V}=I_{N^{2}} \otimes \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} y \tag{22}
\end{equation*}
$$

derived from a discrete version of (11).
The discrete error covariance

$$
\begin{equation*}
P=\mathcal{E}\left\{\tilde{V} \tilde{V}^{T}\right\} \tag{23}
\end{equation*}
$$

can be computed from Equation (21). In Appendix A we show that

$$
\begin{equation*}
\mathcal{E}\left\{\left[\Sigma_{u} U-\Sigma_{w} W\right]\left[\Sigma_{u} U-\Sigma_{w} W\right]^{T}\right\}=\sigma_{u}^{2} \Sigma \tag{24}
\end{equation*}
$$

Since $\Sigma$ is symmetric and, if the problem is well-posed, invertible, solving (21) for $\tilde{V}$, substituting into (23), and taking expected values yields

$$
\begin{align*}
P & =\sigma_{u}^{2} \Sigma^{-1} \Sigma \Sigma^{-1} \\
& =\sigma_{u}^{2} \Sigma^{-1} . \tag{25}
\end{align*}
$$

Thus the error covariance matrix is simply a scalar times the inverse of $\Sigma$, the matrix used to compute the velocity estimates in (22). In light of the rather complicated preceding development the simplicity of this result seems astonishing. It is not so surprising, however, if we look at the problem in a slightly different light. In particular, if (22) is assumed to be the linear minimum mean square error (LMMSE) estimate for a discrete (vector) stochastic estimation problem, then (25) follows immediately by inspection (cf. [38]). In contrast, we found (25) through discretization of the continuous error equations (13). The fact that (22) implies (25) proves that (22) must be the LLMSE estimate for some discrete problem, but to show this fact by deriving an explicit density for the discretized velocity would be at least as complicated as our approach. Indeed, a less careful discretization will not lead to this nice result.

## D. Performance Measure

For typical image sizes the dimensions of $\Sigma-2 N^{2} \times 2 N^{2}-$ make the computation of $\Sigma^{-1}$ impossible. For our purposes, however, all we need is a scalar characterization of the error covariance. A reasonable choice is the velocity error at each pixel averaged across the entire image. Accordingly, since the diagonal elements represent the variances of each component of the velocity at each pixel, we use the scalar

$$
\begin{equation*}
p=\frac{1}{2 N^{2}} \operatorname{tr}[P], \tag{26}
\end{equation*}
$$

as a total measure of error. This quantity can be computed recursively by setting $p^{0}=0$ and using the following iteration

$$
\begin{equation*}
p^{i} \leftarrow p^{i-1}+\sigma_{u}^{2} X_{i}^{i} \quad \text { for } i=1, \ldots, 2 N^{2} \tag{27}
\end{equation*}
$$

where $X^{i}$ is the solution of

$$
\begin{equation*}
\Sigma X^{i}=e_{i} \tag{28}
\end{equation*}
$$

and $e_{i}$ is the $i$ th column of $I_{2 N^{2}}$. The performance is then given by $p=p^{2 N^{2}}$.
Our global measure of optical flow performance, $p$, is obviously a function of both the process noise variance $\sigma_{u}^{2}$ and $\Sigma$. In turn $\Sigma$ is a function of the measurement noise variance $\sigma_{w}^{2}$ and the output gain $\nabla \varphi$. Since the output gain is specified by the brightness function, it is clear that $p$ is a function of the brightness function $\varphi$ and that there might exist a $\varphi$ that will minimize $p$. This is the subject of Section IV. To those who have actually used optical flow, however, it may seem odd that our error does not depend on $v$, the actual velocity. It
turns out that this is due to an inadequacy in our modeling up to this point. In fact, it will be shown in Section IV that modeling errors in the calculation of the temporal derivative of brightness will lead to the expected dependence of the performance on velocity and to the existence of an optimal brightness function.

## IV. BRIGHTNESS FUNCTION OPTIMIZATION

In this section we develop a method to determine brightness functions that yield optimal performance. Specifically, we consider parametrized brightness functions of the form $\varphi(\mathbf{r}, t, \theta)=f(\mathrm{p}(\mathbf{r}, t), \theta)$ where $\theta$ is a parameter vector. Our objective is to determine the parameter vector $\theta^{\circ}$ that gives the optimal performance.

The reasons why an optimal brightness might exist at all are not immediately apparent from the development of Section III. For example, consider the ideal case where the output gain $\nabla \varphi$ is known exactly and the measurement $-\varphi_{t}$ is the exact temporal derivative degraded by additive white Gaussian noise (WGN) with zero mean and variance $\sigma_{w}^{2}$. According to our development, as $\nabla \varphi \rightarrow 0$, the brightness function becomes a constant causing the motion to be unobservable and $p \rightarrow \infty$. In contrast, if $\nabla \varphi \rightarrow \infty$ the signal-to-noise ratio (SNR) increases without bound and $p \rightarrow 0$. It seems therefore that in the ideal case, the optimal brightness function is one where $\nabla \varphi=\infty$. This result, however, ignores two critical factors: first, that the elements of $\nabla \varphi$ cannot simultaneously go to $\infty$ for all $\mathbf{r}$ and second, that in practice the spatial and temporal derivatives are estimated from the data. In this paper we ignore the first issue by considering classes of brightness functions with inherent spatial diversity in their gradients; optimization across these classes is a subject of future research. Instead, in this paper we treat the optimization within a single such class. To do so we model the measurement of the gradient as exact and examine the effects of inexact measurement of the temporal derivative.

The derivative of a function $g(x)$ differentiable on the interval $[a, b]$ can be expressed as [35]

$$
g^{\prime}(x)=\frac{g(x+h)-g(x)}{h}-\frac{h}{2} g^{\prime \prime}(\zeta)
$$

where $\zeta \in[a, b]$. Therefore when $g^{\prime}(x)$ is approximated by the forward difference $[g(x+h)-$ $g(x)] / h$, the approximation error is a function of the sampling interval $h$ and the second derivative of the function. Accordingly, the temporal derivative of brightness, which is considered to be our measurement, has an additional source of noise besides just additive WGN. This temporal derivative approximation error (TDAE) can be modeled as an another additive noise component, which happens to be dependent on the sampling interval and the second derivative (the Hessian) of the brightness function. In this section we first develop a model for the required additional measurement noise which accounts for the TDAE, and then incorporate these results into the error covariance and performance measure $p$ of Section III. We conclude this section by showing the conditions under which, for a given parametrized
brightness class, an optimal parameter exists.

## A. Measurement Noise Model

We begin by deriving the output equation (8) starting from the discrete temporal derivative approximation. Since in practice the reference configuration may be randomly placed relative to the image frame, we will use

$$
\varphi(\mathbf{r}, t, \theta)=f(\mathrm{p}(\mathbf{r}, t)+\phi, \theta)
$$

where $\phi$ is a random vector with known probability distribution. We assume that the brightness function has been corrupted by additive white Gaussian "imaging" noise $w_{a} \sim N\left(0, \sigma_{a}^{2}\right)$. Assuming a forward difference approximation of the temporal derivative, the measurement is

$$
\begin{equation*}
y(\mathbf{r}, t)=-\frac{1}{\Delta t}\left[\varphi(\mathbf{r}, t+\Delta t, \theta)+w_{a}(\mathbf{r}, t+\Delta t)-\varphi(\mathbf{r}, t, \theta)-w_{a}(\mathbf{r}, t)\right] \tag{29}
\end{equation*}
$$

Expanding $\varphi(\mathbf{r}, t+\Delta t)$ in a Taylor series and rearranging terms yields

$$
\begin{equation*}
y(\mathbf{r}, t)=-\varphi_{t}(\mathbf{r}, t, \theta)-\left[\frac{\Delta t}{2} \varphi_{t t}(\mathbf{r}, t, \theta)+\frac{\Delta t^{2}}{3!} \varphi_{t t t}(\mathbf{r}, t, \theta)+\cdots\right]+\frac{w_{a}(\mathbf{r}, t+\Delta t)-w_{a}(\mathbf{r}, t)}{\Delta t} \tag{30}
\end{equation*}
$$

We can replace the additive noise terms with

$$
\begin{equation*}
w^{\prime}=\frac{w_{a}(\mathbf{r}, t+\Delta t)-w_{a}(\mathbf{r}, t)}{\Delta t} \tag{31}
\end{equation*}
$$

and since $w_{a}$ is white, $w^{\prime} \sim N\left(0,2 \sigma_{a}^{2} / \Delta t^{2}\right)$. Substituting (4) into (30) and keeping only the second derivative term yields

$$
\begin{equation*}
y(\mathbf{r}, t)=\nabla \varphi(\mathbf{r}, t, \theta) \cdot v(\mathbf{r}, t)+w^{\prime}(\mathbf{r}, t)-\frac{\Delta t}{2!} \varphi_{t t}(\mathbf{r}, t, \theta) \tag{32}
\end{equation*}
$$

Now,

$$
\begin{align*}
\varphi_{t t}(\mathbf{r}, t, \theta) & =\frac{\partial^{2}}{\partial t^{2}} f(\mathrm{p}(\mathbf{r}, t)+\phi, \theta) \\
& =\mathrm{p}_{t}^{T}(\mathbf{r}, t) \mathcal{H}[f(\mathrm{p}(\mathbf{r}, t)+\phi, \theta)] \mathrm{p}_{t}(\mathbf{r}, t)+\nabla f(\mathrm{p}(\mathbf{r}, t)+\phi, \theta) \cdot \mathrm{p}_{t t}(\mathbf{r}, t) \tag{33}
\end{align*}
$$

where $\mathcal{H}[\cdot]$ denotes the Hessian with respect to the spatial coordinates. Recall that the reference map $p$ and the velocity $v$ are related through equations (1) and (2). Therefore for a general $v$, we cannot compute a closed form expression for p or $\mathrm{p}_{t}$. If we assume, however, that in a neighborhood of $\mathbf{r}$, the motion is approximately a translation, then

$$
\mathrm{p}(\mathbf{r}, t) \approx \mathbf{r}+v(\mathbf{r}) t
$$

and

$$
\begin{equation*}
\varphi_{t t}(\mathbf{r}, t, \theta) \approx v^{T}(\mathbf{r}) \mathcal{H}[f(\mathrm{p}(\mathbf{r}, t)+\phi, \theta)] v(\mathbf{r}) \tag{34}
\end{equation*}
$$

To keep things simple we let $t=0$, which results in

$$
\begin{equation*}
y(\mathbf{r}, t) \approx \nabla \varphi(\mathbf{r}, 0, \theta) \cdot v(\mathbf{r})+w^{\prime}(\mathbf{r}, 0)-\frac{\Delta t}{2} v^{T}(\mathbf{r}) H(\mathbf{r}, \theta) v(\mathbf{r}) \tag{35}
\end{equation*}
$$

where

$$
H(\mathbf{r}, \theta)=\mathcal{H}[f(\mathrm{p}(\mathbf{r}, 0), \theta)]=\mathcal{H}[f(\mathbf{r}+\phi, \theta)]
$$

In subsequent frames $(t \neq 0)$ an estimate of the reference map can in principle be used to compute $H(\mathbf{r}, \theta)$.

The term $v^{T} H v$ in (35) represents the effect of the TDAE on the output equation, and can be incorporated into the measurement equation in (8) by re-defining the measurement noise as

$$
\begin{equation*}
w(\mathbf{r})=w^{\prime}(\mathbf{r})-\frac{\Delta t}{2} v^{T}(\mathbf{r}) H(\mathbf{r}, \theta) v(\mathbf{r}) \tag{36}
\end{equation*}
$$

## B. Performance Measure

Since $p$ depends on $\sigma_{w}^{2}$ and by Equation (36) $w(\mathbf{r})$ depends on $v(\mathbf{r})$, we see that $p$ inevitably depends on the true velocity. This leads to the important conclusion: computation of $p$ requires prior knowledge about $v(\mathbf{r})$. We also note that since this result comes from the observation equation alone it is valid whether or not $v$ is viewed as random (as in our model) or deterministic and unknown (as in Horn and Schunck's formulation). One simple idea which would seem to take advantage of the random formulation is to derive a probability density for $v^{T} H v$ using the state model in (7) and to use this result to derive a new density for $w(\mathbf{r})$. Unfortunately, (7) implies an infinite a priori variance for $v$, which renders the model useless for such calculations. Instead, we assume that an a priori bound on the velocity $v_{\max }(\mathbf{r})=\left[\begin{array}{ll}\mu_{\max }(\mathbf{r}) & \nu_{\max }(\mathbf{r})\end{array}\right]^{T}$ is available, and use

$$
\begin{equation*}
w(\mathbf{r})=w^{\prime}(\mathbf{r})-\frac{\Delta t}{2} v_{\max }^{T}(\mathbf{r}) H(\mathbf{r}, \theta) v_{\max }(\mathbf{r}) . \tag{37}
\end{equation*}
$$

to model the measurement noise. We note that this assumption is a very weak statement of a priori knowledge of $v$, and is generally available in many applications. More detailed knowledge can certainly be employed, and can be expected to generate better estimates of the optimal brightness function.

We assume that the random vector $\phi$ is uncorrelated with $u$ and $w_{a}$ and that

$$
\mathcal{E}_{\phi}\{H(\mathbf{r}, \theta)\}=0 .
$$

With this assumption, the mean of $v^{T} H v$ is

$$
\begin{equation*}
\mathcal{E}\left\{v_{\max }^{T}(\mathbf{r}) H(\mathbf{r}, \theta) v_{\max }(\mathbf{r})\right\}=0 \tag{38}
\end{equation*}
$$

and the variance is

$$
\begin{align*}
\mathcal{E}\left\{\left(v_{\max }^{T}(\mathbf{r}) H(\mathbf{r}, \theta) v_{\max }(\mathbf{r})\right)^{2}\right\}= & H_{x x}^{2} \mu_{\max }^{4}(\mathbf{r})+4 H_{x x} H_{x y} \mu_{\max }^{3}(\mathbf{r}) \nu_{\max }(\mathbf{r}) \\
& +\left(4 H_{x y}^{2}+2 H_{x x} H_{y y}\right) \mu_{\max }^{2}(\mathbf{r}) \nu_{\max }^{2}(\mathbf{r}) \\
& +4 H_{x y} H_{y y} \mu_{\max }(\mathbf{r}) \nu_{\max }^{3}(\mathbf{r})+H_{y y}^{2} \nu_{\max }^{4}(\mathbf{r}) \\
= & \sigma_{v}{ }^{2}(\mathbf{r}, \theta) \tag{39}
\end{align*}
$$

Therefore substituting (39) into (36) yields

$$
\begin{equation*}
\sigma_{w}^{2}(\mathbf{r}, \theta)=\frac{2}{\Delta t^{2}} \sigma_{a}^{2}+\frac{\Delta t^{2}}{4} \sigma_{v}^{2}(\mathbf{r}, \theta) \tag{40}
\end{equation*}
$$

Equation (40) implies that for a given $v_{\max }$ and fixed $\sigma_{a}^{2}$ and $\Delta t$, the measurement noise variance increases with the curvature of the brightness function.

We now wish to use this new model for the observation noise variance in an expression for the performance $p$. In Appendix A we show that

$$
\begin{equation*}
\Sigma=\Sigma_{m}+\Sigma_{\nabla^{2}} \tag{41}
\end{equation*}
$$

where

$$
\begin{equation*}
\Sigma_{m}=I_{N^{2}} \otimes \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \nabla \varphi \nabla \varphi^{T} \tag{42}
\end{equation*}
$$

and $\Sigma_{\nabla^{2}}$ is related to the discrete form of $L^{\dagger} L$. If we were to use (40) in place of $\sigma_{w}^{2}$ in (42), $\Sigma$ would depend on both the parameter $\theta$ and the spatial coordinate $\mathbf{r}$, which is not what we desire. We note, however, that (40) also implies a spatially varying optimal regularization parameter $\alpha^{2}=\sigma_{w}^{2} / \sigma_{u}^{2}$, which is not used in standard optical flow. Therefore, for each parameter $\theta$ we will pick the constant

$$
\begin{equation*}
\sigma_{w}^{2}(\theta)=\max _{\mathbf{r} \in \Omega} \sigma_{w}^{2}(\mathbf{r}, \theta) \tag{43}
\end{equation*}
$$

which leads to the term

$$
\begin{equation*}
\Sigma_{m}(\theta)=I_{N^{2}} \otimes \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}(\theta)} \nabla \varphi(\theta) \nabla \varphi^{T}(\theta) \tag{44}
\end{equation*}
$$

and hence

$$
\begin{equation*}
\Sigma(\theta)=\Sigma_{m}(\theta)+\Sigma_{\nabla^{2}} \tag{45}
\end{equation*}
$$

Defining

$$
\begin{equation*}
P(\theta)=\sigma_{u}^{2} \Sigma^{-1}(\theta) \tag{46}
\end{equation*}
$$

it follows from (26) that

$$
\begin{equation*}
p(\theta)=\frac{1}{2 N^{2}} \operatorname{tr}[P(\theta)] \tag{47}
\end{equation*}
$$

and finally that the optimal parameter vector is

$$
\begin{equation*}
\theta^{o}=\underset{\theta}{\operatorname{argmin}} p(\theta) \tag{48}
\end{equation*}
$$

Since a closed-form expression for $p(\theta)$ cannot be computed in general, $\theta^{\circ}$ must be computed by numerical optimization methods [39].

## C. Existence of $\theta^{\circ}$

We now discuss conditions under which an optimal parameter $\theta^{\circ}$ exists. We will assume for this discussion that $\theta$ is a scalar belonging to the interval $J$; the extension to vector $\theta$ is straightforward but more cumbersome to explain. We assume that $f(\mathbf{r}, \theta)$ is a continuously twice differentiable function of $\theta$ and that $\Sigma(\theta)$ is invertible at each $\theta$. From these facts we conclude that $p(\theta)$ is a continuous function of $\theta$ and finite on $J$. Hence, from elementary real analysis we know that if $J$ is closed $p(\theta)$ assumes a minimum at some point $\theta^{o} \in J$. If $J$ is not closed, however, then $p(\theta)$ may be minimized on the interval or it may attain its minimum at a limit point not in $J$, in which case we say that $\theta^{\circ}$ does not exist. In the special case that $p(\theta)$ is finite somewhere on $J$ and increases without bound as $\theta$ approaches a limit point not in $J$, then $\theta^{\circ}$ is guaranteed to exist. This is the case in the example we study below and use in the experiments of Section V, so it warrants further study.

Of the two terms comprising $\Sigma(\theta)$ in (45) only $\Sigma_{m}(\theta)$ depends on $\theta$. Furthermore, it can be shown that $\Sigma_{\nabla^{2}}$ by itself is not invertible but that in general (and by above assumption) $\Sigma(\theta)=\Sigma_{m}(\theta)+\Sigma_{\nabla^{2}}$ is invertible. Thus as $\Sigma_{m}(\theta) \rightarrow 0, \Sigma(\theta)$ becomes singular. In addition, we show in Appendix B that as $\Sigma_{m}(\theta) \rightarrow 0, p(\theta)=\left(1 / 2 N^{2}\right) \operatorname{tr}\left[\Sigma^{-1}(\theta)\right] \rightarrow \infty$. Therefore we can conclude that $\theta^{\circ}$ is guaranteed to exist for an interval $J$ that is not closed if the condition

$$
\begin{equation*}
\lim _{\theta \rightarrow a} \Sigma_{m}(\theta)=0, \tag{49}
\end{equation*}
$$

is satisfied at each limit point of $J$ not in $J$.
For example, consider the class of brightness functions

$$
\begin{equation*}
f(\mathbf{p}, \theta)=\frac{A}{2}\left(\sin \left(\theta p_{x}\right) \sin \left(\theta p_{y}\right)+1\right) \tag{50}
\end{equation*}
$$

In this case $\theta$ is the spatial frequency, $J=(0, \infty)$, and we want to find the $\theta$ that will result in optimal performance. To see if $\theta^{\circ}$ will exist, we examine the value of $\Sigma_{m}(\theta)$ as $\theta$ approaches 0 and $\infty$ using Equation (44).

After adding a random shift to the position of the pattern, a simple calculation yields the Hessian matrix

$$
\begin{aligned}
H(\mathbf{r}, \theta) & =\mathcal{H}[f(\mathbf{r}+\phi, \theta)] \\
& =\frac{A \theta^{2}}{2}\left[\begin{array}{cc}
-\sin \left(\theta r_{x}+\phi_{x}\right) \sin \left(\theta r_{y}+\phi_{y}\right) & \cos \left(\theta r_{x}+\phi_{x}\right) \cos \left(\theta r_{y}+\phi_{y}\right) \\
\cos \left(\theta r_{x}+\phi_{x}\right) \cos \left(\theta r_{y}+\phi_{y}\right) & -\sin \left(\theta r_{x}+\phi_{x}\right) \sin \left(\theta r_{y}+\phi_{y}\right)
\end{array}\right] .
\end{aligned}
$$

If, for example, $\phi$ has a uniform distribution on any $2 \pi \times 2 \pi$ square then

$$
\mathcal{E}\{H(\mathbf{r}, \theta)\}=0
$$

and, after simplification, Equations (39) and (40) yield

$$
\sigma_{v}^{2}(\mathbf{r}, \theta)=\frac{A^{2}}{16} \theta^{4}\left(\mu_{\max }^{4}(\mathbf{r})+6 \mu_{\max }^{2}(\mathbf{r}) \nu_{\max }^{2}(\mathbf{r})+\nu_{\max }^{4}(\mathbf{r})\right),
$$

and

$$
\sigma_{w}^{2}(\theta)=\frac{2 \sigma_{a}^{2}}{\Delta t^{2}}+\frac{\Delta t^{2} A^{2}}{64} \theta^{4}\left(\mu_{\max }^{4}+6 \mu_{\max }^{2} \nu_{\max }^{2}+\nu_{\max }^{4}\right) .
$$

The output gain outer product is

$$
\begin{align*}
& \nabla \varphi(\theta) \nabla \varphi^{T}(\theta)= \\
& \quad \frac{A^{2} \theta^{2}}{4}\left[\begin{array}{cc}
\cos ^{2}\left(\theta r_{x}+\phi_{x}\right) \sin ^{2}\left(\theta r_{y}+\phi_{y}\right) & \frac{1}{4} \sin \left(2\left[\theta r_{x}+\phi_{x}\right]\right) \sin \left(2\left[\theta r_{y}+\phi_{y}\right]\right) \\
\frac{1}{4} \sin \left(2\left[\theta r_{x}+\phi_{x}\right]\right) \sin \left(2\left[\theta r_{y}+\phi_{y}\right]\right) & \sin ^{2}\left(\theta r_{x}+\phi_{x}\right) \cos ^{2}\left(\theta r_{y}+\phi_{y}\right)
\end{array}\right] . \tag{51}
\end{align*}
$$

Then, since

$$
\lim _{\theta \rightarrow 0} \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}(\theta)} \frac{A^{2} \theta^{2}}{4}=0
$$

and

$$
\lim _{\theta \rightarrow \infty} \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}(\theta)} \frac{A^{2} \theta^{2}}{4}=0
$$

the optimal frequency $\theta^{\circ}$ is guaranteed to exist on $(0, \infty)$. In the next section we will show numerical simulations of this example and comparisons of theoretical and actual performance.

We note that if the amplitude $A$ in Equation (50) is selected as the parameter over which to optimize, then (49) is satisfied at the limit point 0 , but not at $\infty$. Hence, the optimal $A$ is not guaranteed to exist, and in fact it turns out that $p(\theta)$ is minimized at $A=\infty$. For any practical application, this condition implies that if one is constrained to have a fixed frequency pattern one should increase the amplitude (brightness) of the pattern as far as possible to improve the overall optical flow performance.

## V. EXPERIMENTAL RESULTS

In this section we find the optimal spatial frequency $\theta$ for the product-of-sinusoids brightness function given in Equation (50), and we compare this with actual optical flow results. We use the following true velocity field

$$
v(\mathbf{r}, t)=\left[\begin{array}{cc}
-a & -\omega  \tag{52}\\
\omega & -a
\end{array}\right] \mathbf{r}
$$

where $a=\omega=0.02 \sec ^{-1}$. This velocity field is a combination of a counter-clockwise rotation and a contraction about the origin and is shown in Figure 3. Images of the brightness function at time $t=1 \sec$ (not shown) are computed using Equation (50), where the reference map is given by [see Equations (1) and (2)]

$$
\mathrm{p}(\mathbf{r}, t)=e^{-a t}\left[\begin{array}{cc}
\cos (\omega t) & -\sin (\omega t)  \tag{53}\\
\sin (\omega t) & \cos (\omega t)
\end{array}\right] \mathbf{r}
$$

All images displayed and used in this section are $128 \times 128$ pixels, have values in the range [ 0,255 ], and are corrupted by additive white Gaussian noise with mean zero and variance $\sigma_{w}^{2}=0.003125$. The physical distance between pixels is assumed to be 1.0 cm .

Using the numerically efficient local relaxation method described in [40] and modified for optical flow in [31, 32, 24], we computed an SOF velocity estimate using image pairs derived from the above procedure for each of 50 different values of $\theta$. The mean square error (MSE) between the true velocity and the estimated velocity as a function of $\theta$ is shown in Figure 4 using a solid curve. It is clear from the figure that there is an optimal frequency which is approximately $\theta=0.177 \mathrm{rad} / \mathrm{cm}$. The location of this minimum was predicted through calculation of $p(\theta)$, shown using the dotted line in Figure 4. This performance measure was computed using (47), where we assumed $v_{\max }=2.58 \mathrm{~cm} / \mathrm{sec}$ and $\sigma_{u}^{2}=0.05$ $\sec ^{-2}$. Equation (52) was used to compute $v_{\max }$. The choice for $\sigma_{u}^{2}$ was empirical since the velocity field described in (52) is not random. Together these parameters represent our prior knowledge of the true velocity and are necessary inputs to the optimal frequency algorithm. The experimental plot in Figure 4 shows a performance "well" of approximately one-half decade of frequency where SOF is relatively insensitive to the spatial frequency of the brightness function. In practical applications, therefore, one can expect some measure of robustness in the selection of the a priori velocity model parameters.

While the locations of the optimal frequencies of the experimental and theoretical curves match extremely closely in Figure 4, it is disturbing at first to find that the actual level of
performance does not match the predicted level of performance and that the shape of the two curves are quite different. It turns out that these differences are largely the result of the numerical methods used to solve the optical flow equation. In particular, both SOF and the optimal frequency algorithm are required to solve a large system of equations [(22) and (28), respectively], which must be solved iteratively. The matrix $\Sigma(\theta)$ is ill-conditioned at both low and high frequencies but not near the optimal frequency. Therefore, the rate of convergence of the iterative numerical procedure is very slow away from the optimal frequency, and for practical reasons we could not carry the calculations out to completion. Since this procedure is repeated many times in the calculation of the theoretical curve, this curve is affected more that the experimental curve. In simulations on very small images where exact inversion is possible, we have shown that the experimental and theoretical curves match very well.

In Figure 5, we show the actual SOF estimation error separated into a percent average velocity magnitude error given by

$$
\frac{\sum_{i j}\left|\left\|v_{i j}\right\|-\left\|\hat{v}_{i j}\right\|\right|}{\sum_{i j}\left\|v_{i j}\right\|} \times 100 \%
$$

and an average velocity direction error given by

$$
\frac{1}{N^{2} \pi} \sum_{i j}\left|\arccos \frac{v_{i j} \cdot \hat{v}_{i j}}{\left\|v_{i j}\right\|\left\|\hat{v}_{i j}\right\|}\right| .
$$

While the magnitude curve shown in Figure 5a has roughly the same shape as the MSE curve in Figure 4, the direction error curve in Figure 5b shows a much larger range of frequencies over which it is nearly optimal. In fact, the direction error curve is within 5 degrees of optimality over about 1.5 decades of spatial frequency. This result seems to confirm the common observation that the qualitative performance of SOF is better than its quantitative performance. Also note that the minimum direction error occurs at a higher frequency than the minimum magnitude error. This phenomenon suggests first that a qualitative performance measure would result in a different optimal frequency and, second, that in practice it is better to err in favor of higher frequencies when determining the optimal brightness function.

Three examples of SOF velocity estimates are shown in Figures 6b, 6d and 6f superimposed over their respective brightness functions shown at $t=0$. The true velocity field is shown superimposed over the same brightness functions in Figures 6a, 6c and 6e for comparison purposes. Figures 6a and 6 b correspond to a frequency of $\theta=0.05 \mathrm{rad} / \mathrm{cm}$; Figures 6 c and 6 d to $\theta=0.177 \mathrm{rad} / \mathrm{cm}$; and Figures 6 e and 6 f to $\theta=0.5 \mathrm{rad} / \mathrm{cm}$. When the
frequency is too low as in Figure 6b, the SOF estimate tends to exhibit directional errors because there is little variation in the brightness function gradient across the image. This is a good example of the effects of the aperture problem. Because the output equation (4) only provides information about the velocity in the direction of the brightness gradient at each pixel, information about the orthogonal component must be obtained by smoothing. But if there is little local variation in the direction of the gradient, as in this low-frequency image, smoothing will fail to capture the detailed fluctuations of the local velocity field.

In contrast, the high frequency fluctuations of the brightness function in Figure 6 f have plenty of local gradient variations. The poor performance of SOF in this case results from increased effective noise in the calculation of the temporal derivative at each pixel, as described by Equation (40). Here, the larger curvature of the brightness function causes larger effective observation noise, thus requiring a larger regularization parameter $\alpha^{2}=\sigma_{w}^{2} / \sigma_{u}^{2}$. Therefore, at high frequencies the SOF estimate tends to be oversmoothed, and since the velocity mean is zero this typically means that the velocity estimates will be too small. Figure 6d shows the optimal result; comparing this velocity field to the true velocity shown in Figure 6c reveals a nearly perfect result.

## VI. DISCUSSION

We have described a method for finding the optimal brightness function by formulating SOF as an optimal linear smoothing problem and deriving an a priori performance measure based on the estimation error covariance and the effect of the temporal derivative approximation error. Our results show that the performance of SOF is mainly a function of the curvature of the brightness function. At low curvatures, there is little variation in the spatial gradient across the image and the underlying motion is obscured by the aperture problem. At high curvatures, the numerical computation of the temporal derivative increases the measurement noise variance and the resulting optimal estimate is oversmoothed. The optimal brightness function represents the optimal tradeoff between these two effects. This effect of frequency on the performance of SOF suggests that our methods may have applications in the area of multiscale computation of optical flow. Since resolution reduction has the effect of modifying the frequency content of an image, it is possible that our methods may be used to determine the optimal resolution at which to process images using SOF. In addition, the stochastic formulation of SOF presented in this paper has potential applications in the areas of recursive estimation for incremental problems and the derivation of confidence measures.

One interpretation suggested by the results of Section V is that the performance of SOF is determined by the condition of the matrix $\Sigma$. In particular, $\Sigma(\theta)$ is ill-conditioned at low frequencies because the entries of $\nabla \varphi(\theta) \nabla \varphi^{T}(\theta)$ are small [see (51)]. At high frequencies, $\Sigma(\theta)$ is ill-conditioned because the high curvature of the brightness function increases the measurement noise variance causing $\sigma_{u}^{2} / \sigma_{w}^{2}(\theta)$ to be small. Based on this observation, it seems likely that the brightness function yielding the most well-conditioned $\Sigma$ is nearly the same as that which minimizes $p(\theta)$. This relationship between the condition of $\Sigma(\theta)$ and the error covariance makes sense from a stochastic viewpoint also. When $\Sigma(\theta)$ is ill-conditioned, small perturbations in the input (measurements) can result in large perturbations in the estimate which corresponds to a large error covariance. This relationship clearly deserves further investigation.

Although our method depends on the parameters $\sigma_{u}^{2}$ and $v_{\max }$, which often must be determined empirically, our simulation results indicate that SOF is fairly insensitive to these parameters provided that they fall in the proper range. We have also shown that the direction of optical flow velocity estimates is more robust to nonoptimal brightness functions than the magnitude. This result confirms the common observation that the qualitative performance of SOF is better than its quantitative performance.

One disadvantage of our model is that while the calculations required to compute the
performance measure are straight-forward, they are computationally burdensome due to the large dimension of $\Sigma$. Any application requiring real-time or near-real-time modification the brightness pattern, depending on time-varying velocity information for example, would require extensive precalculation to build up tables of optimal patterns which depend on the input parameters. In such cases, further effort to reduce the computational burden (perhaps through examination of the relationship of condition to error) would be fruitful. Another limitation of our approach is that it depends on the a priori specification of a particular parametric class of brightness functions. Although the results of Section IV.C show that the existence of a finite parameter that will provide optimal performance can be shown, we have no idea whether there may exist another class with better overall performance. In future work we therefore plan to examine brightness function optimization over non-parametric functional classes.

At present, the only application of our parameter optimization algorithm is in the area of MR tagging. Other potential applications include estimation of fluid flow where the brightness function may be changed by injecting dyes and biomechanical studies of limb motion where the brightness function may be changed by painting the body. In MR tagging applications, the spatially modulated magnetization (SPAMM) brightness function can be applied to a given tissue using a standard MR scanner [18, 19, 24]. The SPAMM brightness function is similar to the product-of-sinusoids pattern used in this paper and can be parameterized by spatial frequency. In separate simulation experiments we have shown that an optimal spatial frequency exists for SPAMM and the optimization methods presented in this paper predict the optimal frequency with an accuracy comparable to that shown in Section V for a product-of-sinusoids. In tagging applications, however, the brightness function can only be controlled in the initial image. The brightness function in subsequent images will be distorted by the motion of the tissue, and may no longer be optimal. Future work in this area should include the development of a scheme for optimizing the parameter vector over multiple image frames and MR experiments to study the performance and robustness of our algorithm in a practical setting.

## APPENDIX A

In this appendix we prove Equation (24). We begin by defining the matrices $F$ and $\Delta_{F}$ in Equation (20b), then derive explicit equations for the estimate boundary conditions, and finally combine the estimate boundary conditions with the interior equations of (20a) to form the error system of (21). Equation (24) then follows immediately.

## 1) Definition of $F$ and $\Delta_{F}$

The matrices $F$ and $\Delta_{F}$ are required for the discrete version of the continuous velocity field state model of (7). When (7) is discretized on an $N \times N$ lattice as shown in Figure 2 the resulting discrete state model is

$$
L v_{i j}=u_{i j}=\left[\begin{array}{c}
u_{x}  \tag{54}\\
u_{y}
\end{array}\right]_{i j}=\left[\begin{array}{c}
u_{x}^{\mu} \\
u_{x}^{\nu} \\
u_{y}^{\mu} \\
u_{y}^{\nu}
\end{array}\right]_{i j}
$$

for $i j \in \tilde{\Omega}$, and the boundary condition is of the form

$$
\begin{equation*}
F V_{b}=\Delta_{F} U_{b} \tag{55}
\end{equation*}
$$

where $V_{b}$ and $U_{b}$ are boundary vectors as defined in (19).
The boundary condition must be carefully constructed so that the continuous and discrete state models agree. In particular Figure 7 shows a $5 \times 5$ pixel lattice with the nearest neighbor differences shown by the solid and dashed lines. Since the difference operator $L$ is one-sided, it can only specify the differences shown by the dashed lines. Therefore, the boundary condition must specify the remaining differences shown by the solid lines. Accordingly, the top boundary conditions are

$$
\begin{equation*}
\Upsilon_{T} V_{1}+\Delta_{T} V_{2}+\Delta_{1} v_{1}^{\prime}=\Delta_{2} U_{1} \tag{56}
\end{equation*}
$$

where

$$
\begin{gathered}
\Upsilon_{T}=\left[\begin{array}{lllll}
A_{0} & A_{4} & & & \\
& A_{0} & A_{4} & & \\
& & \ddots & & \\
& & & A_{0} & A_{4}
\end{array}\right]_{4(N-1) \times 2 N} \\
\Delta_{T}=\left[\begin{array}{lllll}
0 & 0 & & & \\
& A_{2} & 0 & & \\
& & \ddots & \ddots & \\
& & & A_{2} & 0
\end{array}\right]_{4(N-1) \times 2 N}
\end{gathered}
$$

$$
\begin{aligned}
& \Delta_{1}=\left[\begin{array}{lllll}
A_{2} & 0 & & & \\
& 0 & 0 & & \\
& & \ddots & \ddots & \\
& & & 0 & 0
\end{array}\right]_{4(N-1) \times 2(N-2)} \\
& \Delta_{2}=\left[\begin{array}{lllll}
I_{4} & & & \\
& I_{4} & & & \\
& & \ddots & & \\
& & & I_{4} & 0
\end{array}\right]_{4(N-1) \times 4 N}
\end{aligned}
$$

The left boundary conditions are

$$
\begin{equation*}
\Upsilon_{L} v_{1}^{\prime}+\Delta_{L} v_{2}^{\prime}+\Delta_{3} V_{N}=u_{1}^{\prime} \tag{57}
\end{equation*}
$$

where

$$
\begin{gathered}
\Upsilon_{L}=\left[\begin{array}{lllll}
A_{0} & A_{2} & & & \\
& A_{0} & A_{2} & & \\
& & \ddots & & \\
& & & A_{0} & A_{2} \\
& & & & A_{0}
\end{array}\right]_{4(N-2) \times 2(N-2)} \\
\Delta_{L}=\left[\begin{array}{lllll}
A_{4} & & & \\
& A_{4} & & \\
& & & \ddots & \\
& & & & A_{4}
\end{array}\right]_{4(N-2) \times 2(N-2)} \\
\Delta_{3}=\left[\begin{array}{cccc}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & & & \\
A_{2} & 0 & \cdots & 0
\end{array}\right]_{4(N-2) \times 2 N}
\end{gathered}
$$

The bottom boundary conditions are

$$
\begin{equation*}
\Upsilon_{B} V_{N}=\Delta_{B} U_{N} \tag{58}
\end{equation*}
$$

where

$$
\Upsilon_{B}=\left[\begin{array}{ccccc}
-I & I & & & \\
& -I & I & & \\
& & \ddots & & \\
& & & -I & I
\end{array}\right]_{2(N-1) \times 2 N}
$$

$$
\Delta_{B}=\left[\begin{array}{ccccc}
A_{4}^{T} & 0 & & & \\
& A_{4}^{T} & 0 & & \\
& & \ddots & \ddots & \\
& & & A_{4}^{T} & 0
\end{array}\right]_{2(N-1) \times 4 N}
$$

The right boundary conditions are

$$
\begin{align*}
-e_{N}^{T} V_{1}+e_{1}^{T} v_{N}^{\prime} & =\Delta_{4} U_{1}  \tag{59}\\
\Upsilon_{R} v_{N}^{\prime}+\Delta_{R} V_{N} & =\Delta_{5} u_{N}^{\prime} \tag{60}
\end{align*}
$$

where

$$
\begin{aligned}
& \Upsilon_{R}=\left[\begin{array}{ccccc}
-I & I & & & \\
& -I & I & & \\
& & \ddots & & \\
& & & -I & I \\
& & & & -I
\end{array}\right]_{2(N-2) \times 2(N-2)} \\
& \Delta_{R}=\left[\begin{array}{cccc}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & & & \\
0 & 0 & \cdots & I
\end{array}\right]_{2(N-2) \times 2 N} \\
& e_{1}=\left[\begin{array}{lll}
I & \cdots & \cdots
\end{array}\right]^{T} \\
& e_{N}=[0 \cdots 0 I]^{T} \\
& \Delta_{4}=\left[\begin{array}{llll}
0 & \cdots & 0 & A_{2}^{T}
\end{array}\right]^{T} \\
& \Delta_{5}=I_{N-2} \otimes A_{2}^{T} \text {. }
\end{aligned}
$$

Finally, using (56) - (60) the form of Equation (55) requires that

$$
F=\left[\begin{array}{cccccccc}
\Upsilon_{T} & \Delta_{T} & 0 & 0 & \Delta_{1} & 0 & 0 & 0  \tag{61}\\
0 & 0 & \Upsilon_{B} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \Delta_{3} & 0 & \Upsilon_{L} & \Delta_{L} & 0 & 0 \\
-e_{N}^{T} & 0 & 0 & 0 & 0 & 0 & e_{1}^{T} & 0 \\
0 & 0 & \Delta_{R} & 0 & 0 & 0 & \Upsilon_{R} & 0
\end{array}\right]
$$

and

$$
\Delta_{F}=\left[\begin{array}{cccccccc}
\Delta_{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0  \tag{62}\\
0 & 0 & \Delta_{B} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & I_{4(N-2)} & 0 & 0 & 0 \\
\Delta_{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \Delta_{5} & 0
\end{array}\right] .
$$

## 2) Estimation Error Boundary Conditions

Having found the form of $F$ and $\Delta_{F}$, we can now expand the estimation error boundary conditions of (20b) in terms of the matrices defined in the previous section. The matrix $E$ is given by (cf. [37])

$$
E=\left[\begin{array}{cccccccc}
0 & \Theta_{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
-\Theta_{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \Theta_{2} & 0 & 0 & 0 & 0 \\
0 & 0 & -\Theta_{1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \Theta_{3} & 0 & 0 \\
0 & 0 & 0 & 0 & -\Theta_{4} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \Theta_{4} \\
0 & 0 & 0 & 0 & 0 & 0 & -\Theta_{3} & 0
\end{array}\right]
$$

where

$$
\begin{aligned}
\Theta_{1} & =I_{N}^{0} \otimes A_{1}^{T} \\
\Theta_{2} & =I_{N}^{0} \otimes A_{2}^{T} \\
\Theta_{3} & =I_{N-2} \otimes A_{3}^{T} \\
\Theta_{4} & =I_{N-2} \otimes A_{4}^{T}
\end{aligned}
$$

and

$$
I_{N}^{0}=\left[\begin{array}{lllll}
0 & & & & \\
& 1 & & & \\
& & \ddots & & \\
& & & 1 & \\
& & & & 0
\end{array}\right]_{N \times N}
$$

Performing the matrix multiplications in (20b) and applying the identities

$$
e_{1}^{T} v_{1}^{\prime}=e_{1}^{T} V_{2}
$$

$$
\begin{aligned}
e_{N-2}^{T} v_{1}^{\prime} & =e_{1}^{T} V_{N-1} \\
e_{1}^{T} v_{N}^{\prime} & =e_{N}^{T} V_{2} \\
e_{N-2}^{T} v_{N}^{\prime} & =e_{N}^{T} V_{N-1}
\end{aligned}
$$

derived from duplicate entries in the boundary vectors (see Section III. B.), yields the boundary conditions

$$
\begin{align*}
\left(\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} \Xi_{H}+\Delta_{H}\right) \tilde{V}_{1}-\tilde{V}_{2}= & -\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} W_{1}+\Psi_{0} U_{1}  \tag{63}\\
\left(\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} \Xi_{H}+\Delta_{H}\right) \tilde{V}_{N}-\tilde{V}_{N-1}= & -\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} W_{N}+?_{B} U_{N} \\
& +\Psi_{1} U_{N-1}  \tag{64}\\
\left(\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{V}^{T} \Xi_{V}+\Delta_{V}\right) \tilde{v}_{1}^{\prime}-\tilde{v}_{2}^{\prime}-e_{1} e_{1}^{T} \tilde{V}_{1}-e_{N} e_{1}^{T} \tilde{V}_{N}= & -\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{V}^{T} w_{1}^{\prime}+\Delta_{1}^{T} \Delta_{2} U_{1} \\
& +\Upsilon_{L}^{T} u_{1}^{\prime}  \tag{65}\\
\left(\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{V}^{T} \Xi_{V}+\Delta_{V}\right) \tilde{v}_{N}^{\prime}-\tilde{v}_{N-1}^{\prime}-e_{1} e_{N}^{T} \tilde{V}_{1}-e_{N} e_{N}^{T} \tilde{V}_{N}= & -\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{V}^{T} w_{N}^{\prime}+\Upsilon_{R}^{T} \Delta_{5} u_{N}^{\prime} \\
& +\Theta_{4} u_{N-1}^{\prime}+e_{1} \Delta_{4} U_{1} \tag{66}
\end{align*}
$$

where

$$
\begin{gathered}
\Xi_{H}=I_{N} \otimes C \\
\Xi_{V}=I_{N-2} \otimes C \\
\Delta_{H}=\left[\begin{array}{ccccc}
2 I & -I & & \\
-I & 3 I & -I & \\
& & & \ddots & \\
\Delta_{V}=\left[\begin{array}{ccccc}
3 I & -I & & & \\
-I & 3 I & -I & & \\
\\
& & & & \\
& & & \\
& & -I & & \\
\hline
\end{array}\right]_{2 N \times 2 N} \\
& & & -I & 3 I
\end{array}\right]_{2(N-2) \times 2(N-2)}
\end{gathered}
$$

$$
\begin{aligned}
& ?_{B}=\left[\begin{array}{cccccc}
-A_{4}^{T} & & & & \\
A_{4}^{T} & -A_{4}^{T} & & & \\
& & \ddots & & \\
& & & A_{4}^{T} & -A_{4}^{T} & \\
& & & A_{4}^{T} & 0
\end{array}\right]_{2 N \times 4 N} \\
& \Psi_{0}=\left[\begin{array}{llllll}
A_{0}^{T} & & & & \\
A_{4}^{T} & A_{0}^{T} & & & \\
& & \ddots & & \\
& & A_{4}^{T} & A_{0}^{T} & \\
& & & A_{4}^{T} & -A_{2}^{T}
\end{array}\right]_{2 N \times 4 N}
\end{aligned}
$$

and

$$
\Psi_{1}=I_{N} \otimes A_{2}^{T}
$$

## 3) Total Error System

The interior equations given in (20a) and the left and right boundary conditions given in (65) and (66) can be written in the form

$$
\begin{equation*}
\Phi_{1} \tilde{V}_{i-1}+\left(\Phi_{0}+\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} \Xi_{H}\right) \tilde{V}_{i}+\Phi_{1} \tilde{V}_{i+1}=\Psi_{1} U_{i-1}+\Psi_{0} U_{i}-\frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \Xi_{H}^{T} W_{i} \tag{67}
\end{equation*}
$$

for $2 \leq i \leq N-1$ where

$$
\Phi_{0}=\left[\begin{array}{ccccc}
3 I & -I & & & \\
-I & 4 I & -I & & \\
& & \ddots & & \\
& & -I & 4 I & -I \\
& & & -I & 3 I
\end{array}\right]_{2 N \times 2 N}
$$

and

$$
\Phi_{1}=I_{N} \otimes-I
$$

Equation (67) and the top and bottom boundary conditions given in (63) and (64) can be combined to form the system

$$
\begin{equation*}
\Sigma \tilde{V}=\Sigma_{u} U-\Sigma_{w} W \tag{68}
\end{equation*}
$$

where

$$
\begin{gather*}
\Sigma_{u}=\left[\begin{array}{ccccc}
\Psi_{0} & & & & \\
\Psi_{1} & \Psi_{0} & & & \\
& & \ddots & & \\
& & \Psi_{1} & \Psi_{0} & \\
& & & \Psi_{1} & ?_{B}
\end{array}\right],  \tag{69}\\
 \tag{70}\\
\\
\\
\Sigma_{w}=I_{N^{2}} \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} \otimes C^{T}
\end{gather*}
$$

and the matrix $\Sigma$ can be decomposed into

$$
\begin{equation*}
\Sigma=\Sigma_{m}+\Sigma_{\nabla^{2}} \tag{71}
\end{equation*}
$$

where

$$
\begin{equation*}
\Sigma_{m}=I_{N^{2}} \otimes \frac{\sigma_{u}^{2}}{\sigma_{w}^{2}} C^{T} C \tag{72}
\end{equation*}
$$

and

$$
\Sigma_{\nabla^{2}}=\left[\begin{array}{ccccc}
\Delta_{H} & \Phi_{1} & & &  \tag{73}\\
\Phi_{1} & \Phi_{0} & \Phi_{1} & & \\
& & \ddots & & \\
& & \Phi_{1} & \Phi_{0} & \Phi_{1} \\
& & & \Phi_{1} & \Delta_{H}
\end{array}\right]
$$

Since the random vectors $U$ and $W$ are uncorrelated and

$$
\begin{aligned}
\mathcal{E}\left\{U U^{T}\right\} & =\sigma_{u}^{2} I_{4 N^{2}} \\
\mathcal{E}\left\{W W^{T}\right\} & =\sigma_{w}^{2} I_{N^{2}}
\end{aligned}
$$

it follows after a straightforward calculation that

$$
\begin{align*}
\mathcal{E}\left\{\left[\Sigma_{u} U+\Sigma_{w} W\right]\left[\Sigma_{u} U+\Sigma_{w} W\right]^{T}\right\} & =\sigma_{u}^{2} \Sigma_{u} \Sigma_{u}^{T}+I_{N^{2}} \otimes \frac{\sigma_{u}^{4}}{\sigma_{w}^{2}} C^{T} C \\
& =\sigma_{u}^{2} \Sigma \tag{74}
\end{align*}
$$

## APPENDIX B

In this Appendix, we prove that as $\Sigma_{m} \rightarrow 0, \operatorname{tr}\left[\Sigma^{-1}\right] \rightarrow \infty$. We showed in Section III that $\Sigma^{-1}$ is the covariance matrix of the SOF estimate. Since $\Sigma^{-1}$ is a covariance matrix, we know that $\Sigma^{-1}$ is symmetric and positive definite. It can be shown that $\operatorname{tr}[\cdot]$ is a prenorm on the set of symmetric positive definite matrices. Therefore we can lower bound $\operatorname{tr}\left[\Sigma^{-1}\right]$ by [41]

$$
C_{m}\left\|\Sigma^{-1}\right\|_{2} \leq \operatorname{tr}\left[\Sigma^{-1}\right]
$$

where $C_{m}$ is a positive constant and $\|\cdot\|_{2}$ is the spectral norm. Since $\Sigma$ is also symmetric and positive semidefinite,

$$
\Sigma=U^{T} \Lambda U
$$

where $U$ is a unitary matrix and

$$
\begin{gathered}
\Lambda=\operatorname{diag}\left\{\lambda_{1}, \lambda_{2}, \cdots, \lambda_{2 N^{2}}\right\} \\
\lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{2 N^{2}}
\end{gathered}
$$

Also,

$$
\Sigma^{-1}=U \Lambda^{-1} U^{T}
$$

Since $\left\|\Sigma^{-1}\right\|_{2}=1 / \lambda_{1}$,

$$
\begin{equation*}
\operatorname{tr}\left[\Sigma^{-1}\right] \geq \frac{C_{m}}{\lambda_{1}} \tag{75}
\end{equation*}
$$

We can upper bound $\lambda_{1}$ by [41]

$$
\begin{equation*}
\lambda_{1} \leq \lambda_{1}\left(\Sigma_{\nabla^{2}}\right)+\lambda_{2 N^{2}}\left(\Sigma_{m}\right) \tag{76}
\end{equation*}
$$

Since $\Sigma_{\nabla^{2}}$ is singular $\lambda_{1}\left(\Sigma_{\nabla^{2}}\right)=0$. Therefore using (76) in (75) yields

$$
\operatorname{tr}\left[\Sigma^{-1}\right] \geq \frac{C_{m}}{\lambda_{2 N^{2}}\left(\Sigma_{m}\right)}
$$

Since $\lambda_{2 N^{2}}\left(\Sigma_{m}\right) \rightarrow 0$ as $\Sigma_{m} \rightarrow 0, \operatorname{tr}\left[\Sigma^{-1}\right] \rightarrow \infty$ as $\Sigma_{m} \rightarrow 0$.

## References

[1] B.K.P. Horn and B.G. Schunck. Determining optical flow. Artificial Intelligence, 17:185203, 1981.
[2] B.G. Schunck. Image flow segmentation and estimation by constraint line clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(10):1010-1027, October 1989.
[3] B.G. Schunck. The image flow constraint equation. CVGIP, 35:20-46, 1986.
[4] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis, and error estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(5):451-476, May 1989.
[5] K.Y. Wohn, J. Wu, and R.W. Brockett. A contour-based recovery of image flow: Iterative transformation method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(8):746-760, August 1991.
[6] T.J. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(1):90-99, January 1986.
[7] T.J. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6):168-176, June 1991.
[8] M. Subbarao and A.M. Waxman. Closed form solutions to image flow equations for planar surfaces in motion. CVGIP, 36:208-228, 1986.
[9] A.M. Waxman and K. Wohn. Contour evolution, neighborhood deformation and global image flow: Planar surfaces in motion. Int. J. Robotics Res., 4:95-108, 1985.
[10] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. Proceedings of the IEEE, 76(8):917-935, August 1988.
[11] H.-H. Nagel. Displacement vectors derived from second-order intensity variations in image sequences. CVGIP, 21:85-117, 1983.
[12] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(5):565-593, September 1986.
[13] H.-H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. Artificial Intelligence, 33:299-324, 1987.
[14] H.-H. Nagel. On a constraint equation for the estimation of displacement rates in image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(1):1330, January 1989.
[15] P. Werkhoven, A. Toet, and J.J. Koenderink. Displacement estimates through adaptive affinities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):658663, July 1990.
[16] J. Aisbett. Optical flow with an intensity-weighted smoothing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(5):512-522, May 1989.
[17] E.A. Zerhouni, D.M. Parish, W.J. Rogers, A. Yangand, and E.P. Shapiro. Human heart: tagging with MR imaging - a method for noninvasive assessment of myocardial motion. Radiology, 169:59-63, 1988.
[18] L. Axel and L. Dougherty. MR imaging of motion with spatial modulation of magnetization. Radiology, 171:841-845, 1989.
[19] L. Axel and L. Dougherty. Heart wall motion: improved method of spatial modulation of magnetization for MR imaging. Radiology, 172:349, 1989.
[20] E.R. McVeigh and E.A. Zerhouni. Non-invasive measurement of transmural gradients in myocardial strain with magnetic resonance imaging. Radiology, 180(3):677-683, 1991.
[21] J.L. Prince. Reducing the aperture effect by object tagging in MR imaging. In Proc. of the 1990 Conf. on Information Sciences and Systems, page 443. Princeton University, 1990.
[22] J.L. Prince. Cardiac motion estimation from MR image sequences. In Proc. SPIE's Int'l Symp. on Opt. and Optoelect. Appl. Science and Engr.: Mathematical Imaging, Bellingham, WA, 1990.
[23] J. L. Prince and E. R. McVeigh. Optical flow for tagged MR images. In Proc. of the IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, pages 2441-2444, Toronto, May 1991. IEEE Catalog No. 91CH2977-7.
[24] J.L. Prince and E.R. McVeigh. Motion estimation from tagged MR image sequences. IEEE Transactions on Medical Imaging, 11(2):238—249, June 1992.
[25] E.C. Hildreth. Measurement of Visual Motion. MIT Press, 1984.
[26] T.J. Broida and R. Chellappa. Performance bounds for estimating three-dimensional motion parameters from a sequence of noisy images. J. Opt. Soc. Am, 6(6):879-889, June 1989.
[27] J.K. Kearney and W.B. Thompson. Gradient-based estimation of optical flow with global optimization. In Proc. First Conf. Artificial Intell. Applications, pages 376-380, Denver, CO, 1984.
[28] J.K. Kearney, W.B. Thompson, and D.L. Boley. Optical flow estimation: an error analysis of gradient-based methods with local optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9(2):229-244, March 1987.
[29] E.P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR-91, pages 310-315, Lahaina, Maui, Hawaii, June 1991.
[30] T.M. Chin. Dynamic estimation in computational vision. PhD thesis, Massachusetts Institute of Technology, 1991.
[31] A. Rougee, B. C. Levy, and A. S. Willsky. Optic flow estimation inside a bounded domain. Technical Report Technical Report LIDS-P-1589, MIT Laboratory for Information and Decision Systems, 1986.
[32] A. Rougee, B. C. Levy, and A. S. Willsky. Reconstruction of two-dimensional velocityfields as a linear-estimation problem. In First International Conf. on Computer Vision, pages 646-650, London, England, June 1987.
[33] T.S. Denney Jr. and J.L. Prince. On optimal brightness functions for optical flow. In Proc. of the IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, pages 257-260, San Francisco, CA, May 1992.
[34] M.E. Gurtin. Introduction to Continuum Mechanics. Academic Press, 1981.
[35] R.L. Burden and J.D. Faires. Numerical Analysis. PWS Publishers, third edition, 1985.
[36] M.B. Adams, A.S. Willsky, and B.C. Levy. Linear estimation of boundary value stochastic processes - part I: The role and construction of complementary models. IEEE Transactions on Automatic Control, AC-29(9):803-810, September 1984.
[37] B.C. Levy, M.B. Adams, and A.S. Willsky. Solution and linear estimation of 2-D nearest neighbor models. Proceedings of the IEEE, 78(4):627-641, April 1990.
[38] T. Kailath. Lectures on Weiner and Kalman Filtering. Springer-Verlag, second edition, 1981.
[39] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. Numerical Recipes in C. Cambridge University Press, 1988.
[40] C.-C.J. Kuo, B.C. Levy, and B.R. Musicus. A local relaxation method for solving elliptic pdes on mesh-connected arrays. SIAM J. Sci. Stat. Comput., 8(4):550-573, 1987.
[41] R. A. Horn and C. R. Johnson. Matrix Analysis. Cambridge University Press, 1990.


Figure 1: Motion and Reference Maps.


Figure 2: $N \times N$ discrete lattice.


Figure 3: True Velocity Field.


Figure 4: Theoretical and Actual SOF Performance.



Figure 5: (a) Percent Average Velocity Magnitude Error in cm/sec. (b) Average Velocity Direction Error in degrees.

(b)

Figure 6: (a) $0.05 \mathrm{rad} / \mathrm{cm}$ Brightness Function with True Velocity Field. (b) $0.05 \mathrm{rad} / \mathrm{cm}$ Brightness Function with SOF Velocity Field.


Figure 6: (c) $0.177 \mathrm{rad} / \mathrm{cm}$ Brightness Function with True Velocity Field. (d) $0.177 \mathrm{rad} / \mathrm{cm}$ Brightness Function with SOF Velocity Field.

(f)

Figure 6: (e) $0.5 \mathrm{rad} / \mathrm{cm}$ Brightness Function with True Velocity Field. (f) $0.5 \mathrm{rad} / \mathrm{cm}$ Brightness Function with SOF Velocity Field.


Figure 7: $5 \times 5$ discrete lattice.

## Figure Captions

Figure 1. Motion and reference maps.
Figure 2. $N \times N$ discrete lattice.
Figure 3. True velocity field.
Figure 4. Theoretical and actual SOF performance.
Figure 5. (a) Percent average velocity magnitude error in $\mathrm{cm} / \mathrm{sec}$. (b) Average velocity direction error in degrees.

Figure 6. (a) $0.05 \mathrm{rad} / \mathrm{cm}$ brightness function with true velocity field. (b) $0.05 \mathrm{rad} / \mathrm{cm}$ brightness function with SOF velocity field. (c) $0.177 \mathrm{rad} / \mathrm{cm}$ brightness function with true velocity field. (d) $0.177 \mathrm{rad} / \mathrm{cm}$ brightness function with SOF velocity field. (e) 0.5 $\mathrm{rad} / \mathrm{cm}$ brightness function with true velocity field. (f) $0.5 \mathrm{rad} / \mathrm{cm}$ brightness function with SOF velocity field.

Figure 7. $5 \times 5$ discrete lattice.

