Published in final edited form as:

IEEE Trans Image Process. 2010 January ; 19(1): 1–10. doi:10.1109/TIP.2009.2032341.

Optical Flow Estimation for a Periodic Image Sequence

Ling Li and Yongyi Yang, IEEE [Senior Member]

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA

Abstract

We propose a temporal modeling approach for determining image motion from a sequence of images wherein the inherent motion is periodic over time. To exploit the periodic nature of the motion, we use a Fourier harmonic representation to model the temporal evolution of the motion field for the entire sequence. We then determine the motion field simultaneously for the different image frames by estimating the parameters of this representation model, where the model order in the Fourier representation serves as a regularization parameter on the temporal coherence of the motion field. This approach can take advantage of the statistics of all the available data in the image sequence. In our experiments, we tested the proposed approach on several motion types at different noise levels, including translational motion, convergent/divergent motion, and cardiac motion. Our results demonstrate that this approach could lead to more robust estimation of the motion field in the presence of strong imaging noise compared to a frame-by-frame estimation approach.

Index Terms

Fourier harmonics; motion estimation; optical flow; spatiotemporal regularization

I. Introduction

Motion field modeling and estimation is of considerable interest in a variety of applications in computer vision and image sequence processing. It often constitutes an important, low-level processing step for advanced image analysis tasks. For example, it can be used for detection of object boundary [1], [2] as well as for structural inference of 3-D objects from an image sequence [3], [4]. In image processing, motion-compensated processing has proven to be valuable in problems ranging from image sequence compression [5], [6], noise filtering [7], to image super-resolution [8], [9]. In our recent work [10], [11], [30], we demonstrated the benefit of using motion-compensated processing in spatiotemporal methods for reconstruction of gated images in a cardiac cycle, which was shown to be effective for noise reduction and thereby improving the quality of reconstructed images. A key element in these motion-compensated processing methods is the use of image motion to exploit the similarity among the different image frames in a sequence.

Among the many different methods developed in the literature for motion modeling, the optical flow based approach [12] is perhaps still among the most common in use today. In optical flow modeling, a fundamental assumption is that the image intensity at an object point remains constant along its motion trajectory. Such a condition, however, is well known to be insufficient for unique determination of the motion field, which is known as the aperture problem [13]. As a consequence, additional constraints are often introduced to complement this ill-posed condition [14]. Among them, spatial coherence (i.e., smoothness) on the motion field is the most predominantly used [12], which is based on the assumption that object points in close vicinity of each other tend to move in a similar direction.

Since the original work in [12] and [15], there have been significant interests in development of optical flow based methods for motion estimation. Among these methods some are designed to improve the motion field by introducing spatially adaptive smoothness constraints (e.g., piecewise smooth was used in [16] and [17]), while others aim to improve the robustness of motion field in the presence of noise and outliers [18]. Besides spatial smoothness, there also exists work on assumption of temporal coherence on the motion field [19]–[21], where temporal smoothness constraints were proposed to improve the motion field in a sequence. Specifically, an extension of oriented smoothness constraint was introduced in the temporal domain in [19]; an incremental estimation approach, in which the image plane acceleration is assumed constant over time for an image patch, was formulated in [20]. Furthermore, a discontinuity-preserving spatiotemporal smoothness constraint was proposed in [21].

In this paper, we consider the problem of determining the motion field from a periodic image sequence, i.e., we want to estimate the image motion from a sequence of images wherein the inherent motion is periodic over time. For example, this occurs in cardiac gated imaging, where images are obtained at different phases of the periodic cardiac cycle [22]; another example is in respiratory gated imaging, where the respiratory motion of the chest can also be described by a periodic model [23]. While, in principle, one could adopt a frame-by-frame approach to determine the motion for this problem, we propose a joint estimation approach in which the motion field is estimated simultaneously for the entire sequence, of which the goal is to exploit explicitly the inherent periodicity in image motion over time. Such a joint approach can be potentially advantageous on several aspects. First, it allows for the exploitation of statistics of all the available image data in the presence of noise as opposed to frame by frame, which can lead to a more reliable estimate of the motion. In our approach, we propose to use a Fourier harmonic model to describe the temporal behavior of the motion field; we then determine the motion field by estimating the parameters of this representation model. Besides inherently periodic, this model can also be used to enforce temporal coherence on the motion field over time, thereby serving as a temporal constraint to complement the aperture problem. Moreover, as we explain later, with such a motion model one can easily control the degree of temporal smoothing on the motion field by varying the order of harmonics used.

In our experiments, we demonstrate our proposed motion estimation approach by using several types of image motion at different noise levels: first a translational motion sequence, then a divergent/convergent (zoom in/out) motion sequence, followed by images in gated cardiac single photon emission computed tomography (SPECT) perfusion imaging. Gated SPECT is currently the most frequently ordered test in nuclear medicine for diagnosis and evaluation of coronary artery diseases. It can provide valuable diagnostic information about both myocardial perfusion and heart wall motion. However, they are known to be very noisy due to reduced imaging time per gate frame. These images allow us to demonstrate the robustness of the proposed algorithm in the presence of noise corruption. In addition, we also demonstrate the use of the estimated motion by applying it for motion-compensated noise filtering. Our results demonstrate that the proposed approach could lead to more robust estimation of the motion field in the presence of high imaging noise when compared to a conventional frame by frame approach.

The rest of the paper is organized as follows. In Section II, we first describe our periodic motion representation model, followed by the optimization criterion used for model estimation. In Section III, we discuss the iterative estimation algorithm and related implementation issues. Experimental results are given in Section IV, where our proposed algorithm was tested and compared with the classic optical flow method. Conclusions and

future work are given in Section V. We note that the preliminary development of this work was first presented in [24].

Finally, we point out that in this work the focus will be kept on development of the proposed approach for periodic image sequences. However, the benefit of temporal regularization demonstrated for periodic image sequences should be equally applicable for nonperiodic ones. This possibility is discussed in our future work in Section V.

II. Periodic Image Motion Model

A. Image Data Model

Consider an image sequence $I(x, y, t)$, where t denotes time and (x, y) denotes a spatial location in the image domain D . Let (u, v) denote the displacement vector of point (x, y) from time t to $(t + 1)$. That is

$$I(x, y, t) = I(x + u, y + v, t + 1) \quad (1)$$

where the image intensity at an object point is assumed to be constant along its motion trajectory over time. For notational simplicity, it is noted that in (1) the motion components u and v are implicit functions of both (x, y) and t . Our goal is to determine the motion field (u, v) from a given sequence of images $I(x, y, t)$, $t = 0, 1, \dots, T - 1$.

Of course for this problem, one can resort to a classical solution approach by determining the image motion in a successive, frame by frame fashion. However, such an approach would fail to exploit the potential temporal correlation that may exist in the motion fields among the different image frames. In this work, we seek a joint estimation procedure for the motion across the different frames in the sequence, in which the motion fields (u, v) are determined simultaneously for $t = 0, 1, \dots, T - 1$ by exploiting both spatial and temporal regularizations on them.

In particular, we will consider the case that $I(x, y, t)$ is a periodic sequence, that is, $I(x, y, 0) = I(x, y, T)$. In our approach, we explicitly exploit the inherent periodicity in image motion over time by using basis functions to regulate the temporal behavior of the motion fields. As to be demonstrated later in our experiments, such an approach can be more robust to the image noise, leading to improved estimate of the motion fields.

B. Periodic Motion Model

By applying Fourier series expansion, we model the motion components (u, v) at location (x, y) over time t as

$$u(x, y, t) = \sum_{l=1}^L [a_l(x, y) \cos \omega_l t + b_l(x, y) \sin \omega_l t] \quad v(x, y, t) = \sum_{l=1}^L [c_l(x, y) \cos \omega_l t + d_l(x, y) \sin \omega_l t] \quad (2)$$

where $a_l(x, y)$, $b_l(x, y)$, $c_l(x, y)$, $d_l(x, y)$ are the coefficients associated with harmonic component l , $l = 1, 2, \dots, L$, and L is the order of the harmonic representation, which is determined by the inherent image motion in the sequence.

It is noted that no direct-current (DC) component is included in (2). When included, the DC component would represent the amount of net shift of an object point (x, y) during the entire time period T . Thus, it is assumed in (2) that the net displacement is zero during the period. For example, this is typically the case in a cardiac sequence. However, this component can be easily accommodated in subsequent development if it is known to be none zero in a specific application.

In theory, the model in (2) can be used to represent any arbitrary motion trajectories without sacrificing accuracy as long as the model order L is chosen to be high enough. However, in the case of noisy image data, it can be used as a parameter to regularize the degree of temporal smoothness of the motion field. Thus, it can serve as a trade-off between model accuracy and robustness to data noise. By varying the order of harmonics used in the model, one can achieve different degrees of temporal smoothing on the motion fields. Moreover, it can also allow one to incorporate explicitly a temporal smoothing scheme in a spatially adaptive fashion on the motion field. For example, in a sequence where an object exhibits significantly more motion than the background, one can use a higher order harmonic model for the object region and a lower order model for the background in order to better accommodate the spatially varying nature of the motion fields.

C. Optimization Criterion

To determine the motion fields, we seek a solution to minimize the following weighted objective function

$$E(u, v) = E_1(u, v) + \alpha E_2(u, v) \quad (3)$$

where the two energy terms are defined as

$$E_1(u, v) = \sum_{t=0}^{T-1} \int_D [I(x, y, t) - I(x+u, y+v, t+1)]^2 dx dy \quad (4)$$

and

$$E_2(u, v) = \sum_{t=0}^{T-1} \int_D (|\nabla u|^2 + |\nabla v|^2) dx dy. \quad (5)$$

In (4), the energy term $E_1(u, v)$ is used to enforce the intensity constancy condition in (1) between consecutive frames throughout the whole sequence, and the term $E_2(u, v)$ is used to enforce spatial smoothing on the motion fields by constraining the energy of their spatial gradients for the different time frames. In (3), the parameter α is used to balance these two energy terms.

The objective function $E(u, v)$ in (3) can be viewed as an extension of the classical optical flow of Horn and Schunck [12] to multiple image frames. In (3), spatial smoothing is enforced in a uniform fashion. We note that in more recent work there have been variants of this regularization term in order to better characterize the nature of underlying motion field [25]–[28]. For example, piecewise smoothness constraints can be used to deal with discontinuities in the motion field [16]. For this study, however, the classical uniform smoothness constraint is used in (5), as our main purpose is to demonstrate the effect of the temporal model in (2).

Next, we express the objective function $E(u, v)$ in terms of the harmonic motion model in (2). By applying the first-order Taylor series approximation, we have

$$I(x+u, y+v, t+1) \approx I(x, y, t) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t}. \quad (6)$$

Substituting (6) into (4), we get

$$E_1(u, v) = \sum_{t=0}^{T-1} \int_D \left(\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} \right)^2 dx dy. \quad (7)$$

Next, substituting the harmonic motion model in (2) into (7), we obtain

$$E_1 = \sum_{t=0}^{T-1} \int_D \left[\sum_{l=1}^L (I_{a_l} a_l + I_{b_l} b_l + I_{c_l} c_l + I_{d_l} d_l) + I_t \right]^2 dx dy \quad (8)$$

where $I_{a_l} \triangleq (\|x\| \cos \mu t)$, $I_{b_l} \triangleq (\|x\| \sin \mu t)$, $I_{c_l} \triangleq (\|y\| \cos \mu t)$, and $I_{d_l} \triangleq (\|y\| \sin \mu t)$.

Similarly, substituting (2) into (5) and invoking the orthogonality of the harmonic basis functions, we can rewrite the energy term $E_2(u, v)$ as

$$E_2 = \frac{T}{2} \sum_{l=1}^L \int_D (|\nabla a_l|^2 + |\nabla b_l|^2 + |\nabla c_l|^2 + |\nabla d_l|^2) dx dy. \quad (9)$$

The problem now becomes minimization of the objective function in (3) with respect to the unknown coefficients associated with the different orders of harmonic basis functions.

From calculus of variations, the minimization of the objective function in (3) with respect to the unknown parameters $a_l(x, y)$, $b_l(x, y)$, $c_l(x, y)$, $d_l(x, y)$ yields the following Euler–Lagrange equations [12], [18] for $l = 1, 2, \dots, L$

$$\begin{aligned} \beta \nabla^2 a_l - \sum_{t=0}^{T-1} I_{a_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) + I_t \right] &= 0 \beta \nabla^2 b_l \\ - \sum_{t=0}^{T-1} I_{b_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) + I_t \right] &= 0 \beta \nabla^2 c_l - \sum_{t=0}^{T-1} I_{c_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) + I_t \right] \\ &= 0 \beta \nabla^2 d_l - \sum_{t=0}^{T-1} I_{d_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) + I_t \right] = 0 \end{aligned} \quad (10)$$

where $\beta = T/2$, and ∇^2 denotes the Laplacian operator.

III. Implementation Issues

A. Iterative Estimation Algorithm

For numerical implementation, the Laplacian operator in (10) is approximated by a discrete operator, which at a pixel location is equivalent to subtracting the pixel value from a

weighted average of its surrounding neighboring pixels. Therefore, (10) can be written in discrete form as

$$\begin{aligned}
 \beta(a_l - \bar{a}_l) + \sum_{t=0}^{T-1} I_{a_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) \right] &= - \sum_{t=0}^{T-1} I_{a_l} I_t \\
 \beta(b_l - \bar{b}_l) + \sum_{t=0}^{T-1} I_{b_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) \right] &= - \sum_{t=0}^{T-1} I_{b_l} I_t \\
 \beta(c_l - \bar{c}_l) + \sum_{t=0}^{T-1} I_{c_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) \right] &= - \sum_{t=0}^{T-1} I_{c_l} I_t \\
 \beta(d_l - \bar{d}_l) + \sum_{t=0}^{T-1} I_{d_l} \left[\sum_{k=1}^L (I_{a_k} a_k + I_{b_k} b_k + I_{c_k} c_k + I_{d_k} d_k) \right] &= - \sum_{t=0}^{T-1} I_{d_l} I_t
 \end{aligned} \tag{11}$$

where $\bar{a}_l$ denotes a weighted average surrounding a_l such that $\bar{a}_l \approx \frac{1}{L} \sum_{k=1}^L a_k$, and similarly for b_l, c_l, d_l .

As can be seen, (11) is simply a system of linear equations in terms of the unknown coefficients $a_l, b_l, c_l, d_l, l = 1, 2, \dots, L$. Moreover, it can be shown that the coefficient matrix of this system of equations is symmetric and positive definite (Appendix A). In our experiments, the iterative Jacobi method was used for solution of (11). In this method, the unknowns a_l, b_l, c_l, d_l are updated pixel by pixel with the following iteration [see (12), shown at the bottom of the page].

Note that for computational saving the different summation terms over the temporal index t in (12) can be precalculated prior to the numerical iterations, because they are constant with respect to the iteration index n . In our experiments, the iteration in (12) was initialized with zero values.

B. Effect of Spatiotemporal Regularization

There are two parameters associated with the iterative algorithm in (12), namely, the motion model order L and spatial regularization parameter β . These two parameters are used to regulate the resulting motion field both spatially and temporally. Specifically, as in the case of classical optical flow estimation, the parameter β plays the role of balancing between the optical flow constraint term $E_1(u, v)$ and the spatial coherence constraint term $E_2(u, v)$. Its effect can be seen from the iteration algorithm in (12). Consider the situation that in a region the image intensity is relatively uniform. In this case, the derivative terms will be close to zero. When this happens, the parameter β will help stabilize the denominator terms in (12) and consequently avoid the ambiguity associated with the numerical difficulty. Moreover, in the case that the image is subject to noise, the estimates of the partial derivatives from the image data will become noisy as well. In such a case, the parameter β will help mitigate the noise effect of the derivative terms in the denominator terms in (12). On the other hand, it is also clear that when the image data are nearly free of noise, (i.e., when reliable estimates of derivative terms are available), a smaller value should be used for the parameter β so that the image features (reflected by the derivative terms) will play a more dominant role.

In contrast, the model order L is used to regulate the temporal coherence of the motion field. On the surface, a higher order model can be more desirable, because it can offer more degrees of freedom (and, hence, potentially better representation accuracy) than a lower order model. However, its benefit can become diminished with increased noise in the image data. A lower order model in such a case could offer robustness against the noise by exploiting the temporal continuity of the image motion.

$$\begin{aligned}
a_l^{(n+1)} &= \frac{\beta}{\beta + \sum_{t=0}^{T-1} I_{a_l}^2} \bar{a}_l^{(n)} - \frac{\sum_{k \neq l} \left(\sum_{t=0}^{T-1} I_{a_k} I_{a_l} \right) a_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{b_k} I_{a_l} \right) b_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{c_k} I_{a_l} \right) c_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{d_k} I_{a_l} \right) d_k^{(n)} + \sum_{t=0}^{T-1} I_{a_l} I_t}{\beta + \sum_{t=0}^{T-1} I_{a_l}^2} \\
b_l^{(n+1)} &= \frac{\beta}{\beta + \sum_{t=0}^{T-1} I_{b_l}^2} \bar{b}_l^{(n)} - \frac{\sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{a_k} I_{b_l} \right) a_k^{(n)} + \sum_{k \neq l} \left(\sum_{t=0}^{T-1} I_{b_k} I_{b_l} \right) b_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{c_k} I_{b_l} \right) c_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{d_k} I_{b_l} \right) d_k^{(n)} + \sum_{t=0}^{T-1} I_{b_l} I_t}{\beta + \sum_{t=0}^{T-1} I_{b_l}^2} \\
c_l^{(n+1)} &= \frac{\beta}{\beta + \sum_{t=0}^{T-1} I_{c_l}^2} \bar{c}_l^{(n)} - \frac{\sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{a_k} I_{c_l} \right) a_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{b_k} I_{c_l} \right) b_k^{(n)} + \sum_{k \neq l} \left(\sum_{t=0}^{T-1} I_{c_k} I_{c_l} \right) c_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{d_k} I_{c_l} \right) d_k^{(n)} + \sum_{t=0}^{T-1} I_{c_l} I_t}{\beta + \sum_{t=0}^{T-1} I_{c_l}^2} \\
d_l^{(n+1)} &= \frac{\beta}{\beta + \sum_{t=0}^{T-1} I_{d_l}^2} \bar{d}_l^{(n)} - \frac{\sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{a_k} I_{d_l} \right) a_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{b_k} I_{d_l} \right) b_k^{(n)} + \sum_{k=1}^L \left(\sum_{t=0}^{T-1} I_{c_k} I_{d_l} \right) c_k^{(n)} + \sum_{k \neq l} \left(\sum_{t=0}^{T-1} I_{d_k} I_{d_l} \right) d_k^{(n)} + \sum_{t=0}^{T-1} I_{d_l} I_t}{\beta + \sum_{t=0}^{T-1} I_{d_l}^2}
\end{aligned} \tag{12}$$

In our experiments, we will present empirical results to demonstrate the effect of both regularization parameters and L with different image noise levels and different types of image motion.

C. Estimation of Derivative Terms

The numerical iteration in (12) requires the knowledge of the spatial and temporal partial derivatives of the image function $I(x, y, t)$. In our experiments, these partial derivatives (i.e., $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, and $\frac{\partial I}{\partial t}$) were simply estimated by using the first-order forward difference. Since derivatives are known to be very sensitive to the noise in the image owing to their highpass nature, presmoothing was first applied to the image to mitigate the noise effect. The parameters of the presmoothing filter used are given in the experiments in the next section.

D. Local Averaging for the Laplacian Operator

As mentioned above, the Laplacian operator can be approximated as a difference of a pixel from a local average of its surrounding neighbors, i.e., $\nabla^2 a_l \approx \bar{a}_l - a_l$. In our experiments, an 8-neighbor Laplacian operator was used, for which the local average $\bar{a}_l$ at pixel location (i, j) is computed as

$$\bar{a}_l(i, j) = \frac{1}{6} [a_l(i-1, j) + a_l(i, j-1) + a_l(i, j+1) + a_l(i+1, j)] + \frac{1}{12} [a_l(i-1, j-1) + a_l(i-1, j+1) + a_l(i+1, j-1) + a_l(i+1, j+1)]. \tag{13}$$

IV. Numerical Results

In this section, we present numerical results to demonstrate the performance of the proposed motion estimation approach. Specifically, we present three sets of experiments which are designed to test the algorithm for different types of motion: 1) translational motion (Section IV-B), which is typically associated with object displacement or camera panning, 2) convergent/divergent motion (Section IV-C), which is typically associated with camera zooming, and 3) cardiac motion (Section IV-D). This allows us to demonstrate how the proposed approach would perform under different degrees of complexity in motion.

Moreover, to demonstrate the robustness of the algorithm, each experiment is tested with different levels of image noise.

A. Motion Estimation Methods and Evaluation Criteria

In each experiment, the proposed motion estimation method was applied to determine the image motion. This approach yields the frame-to-frame image motion simultaneously for a sequence (referred to as OFE-DFT for convenience). For comparison, we also tested the classical optical flow algorithm of Horn and Schunck [12]. With this approach the image motion in a sequence was estimated in a frame-by-frame fashion (referred to as OFE hereafter).

To quantify the accuracy of the estimated motion, we compare it against the known motion when it is available (the first two experiments). In addition, we also test the estimated motion by computing the frame-to-frame prediction error of the sequence, which is defined as follows:

$$E_{MC} = \frac{1}{T} \sum_{t=0}^{T-1} 10 \log_{10} \left(\frac{\|\mathbf{f}_t\|^2}{\|\mathbf{f}_t - \hat{\mathbf{f}}_{t \leftarrow t+1}\|^2} \right) \text{dB} \quad (14)$$

where $\mathbf{f}_t$ denotes frame t in the sequence, and $\mathbf{f}_{t \leftarrow t+1}$ denotes its motion-compensated prediction from its neighboring frame $\mathbf{f}_{t+1}$. In computing $\mathbf{f}_{t \leftarrow t+1}$, the estimated image motion from frame t to frame $t+1$ is used. For fractional-pixel motion vectors, bilinear interpolation from four nearest integer pixels is used.

Note that the prediction error E_{MC} in (14) is defined on *noiseless* images. In our experiments, the motion will be estimated from noise corrupted images, but the estimated motion will be tested on the noiseless images when computing E_{MC} . This would allow us to evaluate how the estimated motion would conform with the underlying noiseless image data.

B. Experiment 1: Translational Motion

In this experiment, we tested the motion estimation algorithm using an image sequence with translational motion. The image sequence consisted of an image pattern, [simulated using a circular Gaussian shape with $\sigma = 2$, of which the support is limited to the full image domain, as shown in Fig. 1(a)], which was displaced periodically among successive frames. The sequence consisted of 10 frames; during the first half of the sequence the image pattern was translated along the direction $(1, 1)^T$ with a step-size of 0.5 pixels between two consecutive frames, while during the second half it was translated backward with the same step-size toward its starting position. In Fig. 2, the motion fields are shown for two selected frames (#3) of the sequence (the motion vectors are amplified by a factor of 1.5 for clarity in the plot). To test the robustness of the motion estimation algorithm, white Gaussian noise at different levels (SNR = 13.68, 7.72, and 1.63 dB, respectively) was added to the images. The resulting noisy images were then used for subsequent processing. As an example, in Figs. 1(b), we show the noisy images at noise level SNR = 7.72 dB. For brevity, only half of the image frames (odd-numbered) are shown. The image size was 32×32 pixels for each frame. To mitigate the noise effect, the noisy images were first prefiltered with an FIR lowpass filter (length = 12, bandwidth = 0.3, 0.225, 0.15 cycles/pixel respectively for the three noise levels); subsequently, the partial derivative terms were estimated.

To demonstrate the effect of the motion model on the motion field, we tested the motion estimation algorithm for different values of the order L . The results are summarized in Fig. 3 for the case of noise level at SNR = 7.72 dB, in which the motion-compensated prediction error E_{MC} obtained with the estimated motion is shown for different parametric settings;

each curve in the plot was obtained for a fixed model order L by varying the spatial smoothing parameter σ . For this case, the best E_{MC} was obtained with $L = 3$ and $\sigma = 10$.

Similarly, we also tested the algorithm for the other two noise levels and essentially similar results were obtained. For brevity, we summarize these results in Table I, where the obtained best E_{MC} results are listed for different model orders and noise levels; for comparison, the best results are also given for the classical OFE method, for which the images were prefiltered the same way as described above for the proposed OFE-DFT. These results show that the best results were obtained with model order $L = 3$. A higher order model ($L = 4$) did not seem to improve the results, indicating that its benefit simply became outweighed by the noise in the image data. Furthermore, it is interesting to note that the improvement in E_{MC} over the classical OFE is the largest when the noise level is at 1.63 dB.

In Fig. 4, we show the obtained motion fields for two representative frames (#3); the known motion fields for these two frames were shown earlier in Fig. 2. For comparison, we also show in Fig. 4 the corresponding results obtained by the classical OFE method. These results were optimized over the spatial smoothing parameter of the OFE.

From Fig. 4, we observe that the accuracy of the estimated motion fields degrades with the increased noise level. Compared to the classical OFE, the proposed motion model shows good robustness in the presence of strong noise (SNR = 1.63 dB). These results are consistent with the prediction error given earlier in Table I.

C. Experiment 2: Convergent/Divergent Motion

In this experiment, we tested the motion estimation algorithm using an image sequence with convergent/divergent motion. The image sequence was simulated with the same image pattern as in Experiment 1, which was expanding/shrinking in size in successive frames. It consisted of ten frames; during the first half of the sequence the image pattern was expanding (σ was increased by 5% between two consecutive frames), while during the second half it was shrinking successively with the same proportion towards its initial size. In Fig. 5(a) we show the five odd-numbered frames during both the expanding/shrinking phases of the sequence. In Fig. 6, the motion fields are shown for two selected frames (#3) of the sequence (the motion vectors are amplified by a factor of 2 for clarity in the plot). To test the robustness of the motion estimation algorithm, white Gaussian noise at different levels (SNR = 12.13, 6.18, and 1.05 dB, respectively) was added to the images. As an example, in Fig. 5(b), we show the noisy images at noise level SNR = 1.05 dB. The image size was 32×32 pixels. To mitigate the noise effect, the noisy images were prefiltered as in Experiment 1.

In Fig. 7, we show a plot of the motion-compensated prediction error E_{MC} obtained with the estimated motion for the case of noise level at SNR = 6.18 dB; each curve in the plot was obtained for a fixed model order L by varying the spatial smoothing parameter σ . As can be seen, the best E_{MC} was obtained with $L = 3$ and $\sigma = 0.5$.

We also tested the algorithm for the other noise levels, and similar results were obtained. We summarize these results in Table II, where the best E_{MC} results are listed for different model orders and noise levels. It is observed that the best results were obtained with $L = 3$ for the first two noise levels; however, at high noise level SNR = 1.05 dB, $L = 2$ yielded the best results. This demonstrates that when the noise is high, a more restrictive temporal model is more effective for completing the image data.

In Fig. 8, we show the obtained motion fields for two representative frames (#3); the known motion fields of these two frames were shown earlier in Fig. 6. For comparison, in

Fig. 8, we also show the obtained motion field for these two frames by the classical OFE method; these results were optimized over the spatial smoothing parameter of the OFE.

From Fig. 8, we observe that the classical OFE degraded rather quickly with the increase of image noise. We believe that this is largely due to the complexity of the underlying motion field (compared to the simple translational motion in Experiment 1); the algorithm in this case simply failed to recover the details of the motion field. In contrast, the proposed motion model showed much better resilience to noise.

D. Experiment 3: Cardiac Motion

In this experiment, the proposed motion estimation approach was tested using gated cardiac image sequences in single photon emission computed tomography (SPECT), which is an important diagnostic imaging technique currently in use for diagnosis and evaluation of cardiac diseases. In gated cardiac SPECT, the data acquisition is synchronized to the electrocardiogram (ECG) signal, based on which the cardiac cycle is divided into a number of intervals (typically 8 to 16), and a sequence of images is reconstructed for the cardiac cycle. When displayed in cine, this image sequence can provide valuable diagnostic information about the wall motion of the left ventricle [29]. While extremely useful, gated cardiac SPECT images are also known to suffer from noise (e.g., Fig. 9). Our goal is to demonstrate the robustness of the proposed approach in the presence of such noise. In addition, we also demonstrate the utility of the obtained motion for motion-compensated filtering in these images.

In our experiment, images were obtained from simulated gated SPECT imaging with the 4D NURBS-based cardiac-torso (NCAT) 2.0 phantom [30]. The 4D NCAT phantom was based on heart anatomy and cardiac motion derived from real patients. Use of simulated imaging provides the ground truth of the images for evaluation. In our experiment, a total of 16 gated image frames were used for the cardiac cycle. Shown in Fig. 9(a) are six selected frames (#1, #3, #7, #10, #12, #15) of the myocardium (in short axis view) which were obtained from the ideal case of noiseless acquisition. In these images, the image intensity represents the blood perfusion in the left ventricle (which is of primary interest). Shown in Fig. 9(b) are these frames obtained at a noise level that is typical of a clinical acquisition. With the noiseless images as the reference, the average SNR of the 16 frames is approximately 10.06 dB.

The proposed OFE-DFT algorithm was applied to estimate the frame-to-frame motion from the noisy images. To suppress the noise, these images were first prefiltered with a lowpass FIR filter (length 11, bandwidth 0.3 cycles/pixel). We summarize the results in Fig. 10, in which the motion-compensated prediction error E_{MC} is shown for different parametric settings. As can be seen, the best result was obtained when $L = 2$ and $\alpha = 4$.

In Fig. 11, we show the obtained motion fields for four selected frames (#1, #3, #10, #12). For comparison, in Fig. 11 we also show the results obtained by the classical OFE method (second row), which were also optimized based on E_{MC} . As reference, the motion estimated from the noiseless images in Fig. 9 with the OFE algorithm is shown (first row). For clarity, in these plots the motion vectors are amplified by a factor of 2 and shown for only those pixels on the heart wall. One can see that the estimated motion from proposed OFE-DFT is notably more robust to the noise.

Finally, to demonstrate the utility of the estimated image motion, we also tested the estimated motion in the context of motion-compensated filtering. For this purpose, we implemented the following recursive scheme:

$$\mathbf{f}_t^{(n+1)} = \gamma \mathbf{f}_t^{(n)} + (1 - \gamma) \hat{\mathbf{f}}_{t \leftarrow t+1}^{(n)}, k=0, \dots, L-1 \quad (15)$$

where n is the recursive index, and $\hat{\mathbf{f}}_{k \leftarrow k+1}^{(n)}$ denotes the motion-compensated prediction of frame k from frame $k+1$ at the current recursion. During each recursion, the operation in (15) is carried out successively over all the frames in the sequence. In our experiments, $\gamma = 0.5$ was used. We applied motion-compensated filtering to the noisy images in Fig. 9(b). We show in Fig. 9(c) the filtered images based motion from OFE-DFT, where a total of 3 recursions were used; the average SNR of the 16 filtered frames is 14.56 dB (compared to 13.70 dB for classical OFE; the SNR is 10.06 dB for the noisy images). These results demonstrate that the obtained motion from OFE-DFT can be more effective for noise filtering than that from the classical OFE.

V. Conclusions and Future Work

In this paper, we investigated the use of a temporal model for motion estimation in a periodic sequence of images, in which Fourier harmonic functions are used to represent the periodic image motion in the sequence. With this harmonic representation model, the motion field was estimated in a simultaneous fashion for the whole sequence instead of frame by frame by making use of all the available image data. Our results using several motion types demonstrated that this joint estimation approach could yield more robust motion estimation than a frame-by-frame approach in the presence of high imaging noise, which could lead to improved image quality for motion-compensated noise filtering.

Encouraged by the promising results in this paper, in the future, we plan to apply the proposed method to motion-compensated cardiac image reconstruction where the image quality is often hampered by strong imaging noise and cardiac motion. In addition, while in this work the focus has been on periodic image sequences, the benefits of using temporal regularization on the motion field should be equally applicable for nonperiodic sequences. One possibility is to extend our proposed approach to nonperiodic sequences by duplicating the image frames a sequence in a backward order to create a periodic sequence. This will be explored in our future work.

Acknowledgments

This work was supported by the NIH/NHLBI Grant HL65425.

Appendix A

Demonstration of Positive Definite Property of Coefficient Matrix

Below, we demonstrate that the coefficient matrix of the system of linear equations in (11) is strictly positive definite. For simplicity, let vector $\mathbf{w}$ denote the collection of the unknown parameters a_l, b_l, c_l, d_l for $l = 1, 2, \dots, L$. Then the quadratic energy term E_1 in (8) can be written in the form of $E_1 = \mathbf{P}\mathbf{w} + \mathbf{c}^2$, where the operator $\mathbf{P}$ corresponds to the coefficient terms of a_l, b_l, c_l, d_l in (8) and $\mathbf{c}$ is a constant vector; similarly, the energy term E_2 can be written as $E_2 = \mathbf{Q}\mathbf{w}^2$, where the operator $\mathbf{Q}$ corresponds to the gradient terms in (9). Then the objective function E in (3) can be written as $E = \mathbf{P}\mathbf{w} + \mathbf{c}^2 + \mathbf{Q}\mathbf{w}^2$. Therefore, the Euler-Lagrange equations in (10) can be rewritten as $(\mathbf{P}^T\mathbf{P} + \mathbf{Q}^T\mathbf{Q})\mathbf{w} = -\mathbf{P}^T\mathbf{c}$. Clearly, the coefficient matrix $\mathbf{a} \triangleq \mathbf{P}^T\mathbf{P} + \mathbf{Q}^T\mathbf{Q}$ is symmetric; moreover, it is strictly positive definite for $\gamma > 0$ as long as the operators $\mathbf{P}$ and $\mathbf{Q}$ do not have a common eigenvector with which the associated eigenvalue is zero.

References

1. Meygret, A.; Thonnat, M. Segmentation of optical flow and 3D data for the interpretation of mobile objects; Proc. 3rd Int. Conf. Comput. Vision; 1990. p. 238-245.
2. Heitz F, Perez P, Memin E, Bouthemy P. Parallel visual motion analysis using multiscale Markov random fields. Proc. Workshop Visual Motion. 1991:30-35.
3. Buxton BF, Buxton H. Monocular depth perception from optical flow by space time signal processing. Proc. Roy. Soc. London B. 1983; vol. 218:27-47. [PubMed: 6135210]
4. Longuet-Higgins HC, Prazdny K. The interpretation of a moving retinal image. Proc. Roy. Soc. London B. 1980; vol. 208:385-397. [PubMed: 6106198]
5. Yan P, Kassin A. Lossless and near-lossless motion-compensated 4D medical image compression. Proc. IEEE Int. Workshop on Biomedical Circuits and Systems. 2004:13-16.
6. Park, D.; Cho, H.; Cho, S.; Lee, J. A fast motion estimation algorithm for SAD optimization in sub-pixel; Proc. IEEE Int. Symp. Integrated Circuits; 2007. p. 528-531.
7. Biloslavo M, Ramponi G, Olivieri S, Albani L. Joint Kalman-based noise filtering and motion compensated video coding for low bit rate video conferencing. Proc. IEEE Int. Conf. Image Processing. 2000; vol. 1:992-995.
8. Vazquez C, Aly H, Dubois E, Mitiche A. Motion compensated super-resolution of video by level sets evolution. Proc. IEEE Int. Conf. Image Processing. 2004; vol. 3:1767-1770.
9. Le HV, Seetharaman G. A super-resolution imaging method based on dense subpixel-accurate motion fields. J. VLSI Signal Process. Syst. 2006; vol. 42:79-89.
10. Jin M, Yang Y, Wernick MN. Dynamic image reconstruction using temporally adaptive regularization for emission tomography. Proc. IEEE Int. Conf. Image Processing. 2007; vol. 4:141-144.
11. Gravier E, Yang Y, Jin M. Tomographic reconstruction of dynamic cardiac image sequences. IEEE Trans. Image Process. 2007; vol. 16:932-942. [PubMed: 17405427]
12. Horn BKP, Schunck BG. Determining optical flow. Artif. Intell. 1981; vol. 17:185-203.
13. Mann R, Langer MS. Optical snow and the aperture problem. Proc. 16th Int. Conf. Pattern Recognition. 2002; vol. 4:40264.
14. Chin TM, Karl WC, Willsky AS. Probabilistic and sequential computation of optical flow using temporal coherence. IEEE Trans. Image Process. 1994; vol. 3:773-788. [PubMed: 18296246]
15. Lucas, B.; Kanade, T. An iterative image registration technique with an application to stereo vision. Proc. 7th Int. Joint Conf. Art. Intell.; Vancouver, BC, Canada. 1981. p. 674-679.
16. Nagel H-H, Enkelmann W. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Trans. Pattern Anal. Mach. Intell. 1986; vol. 8:565-593. [PubMed: 21869357]
17. Weickert J, Schnorr C. A theoretical framework for convex regularizers in PDE-based computation of image motion. Int. J. Comput. Vis. 2001; vol. 45:245-264.
18. Black MJ, Anandan P. The robust estimation of multiple motions: Parametric and piecewise smooth flow fields. Comput. Vis. Image Understand. 1996; vol. 63:75-104.
19. Nagel, H-H. Proc. Computer Vision. Vol. vol. 427. Berlin, Germany: Springer; 1990. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow; p. 139-148.
20. Black, MJ.; Anandan, P. Robust dynamic motion estimation over time. Proc. IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition; Maui, HI. 1991. p. 292-302.
21. Brox T, Bruhn A, Papenberg N, Weickert J. High accuracy optical flow estimation based on a theory for warping. Proc. 8th Eur. Conf. Computer Vision. 2004; vol. 4:25-36.
22. Beller GA, Zaret BL. Contributions of nuclear cardiology to diagnosis and prognosis of patients with coronary artery disease. J. Amer. Heart Assoc., Circulation. 2000; vol. 101:1465-1478.
23. Olbrich B, Traub J, Wiesner S, Wichert A, Feussner H, Navab N. Respiratory motion analysis: Towards gated augmentation of the liver. Proc. Int. Congress Series. 2005; vol. 1281:248-253.
24. Li, L.; Yang, Y. Optical flow estimation for a periodic image sequence; Proc. IEEE Int. Conf. Image Processing; 2008. p. 833-836.

25. Vemuri BC, Ye J, Chen Y, Leonard CM. Image registration via level-set motion: Applications to atlas-based segmentation. *Med. Image Anal.* 2003; vol. 7:1–20. [PubMed: 12467719]
26. McClelland J, Blackall J, Tarte S, Hughes S, Hawkes D. Non-rigid registration based respiratory motion models of the lung using two parameters. *Med. Phys.* 2007; vol. 34:2516.
27. Ehrhardt J, Werner R, Saring D, Frenzel T, Lu W, Low D, Handels H. An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. *Med. Phys.* 2007; vol. 34:711–721. [PubMed: 17388189]
28. Bereziat D, Herlin I, Younes L. A generalized optical flow constraint and its physical interpretation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition.* 2000; vol. 2:2487.
29. Garcia EG. Imaging guidelines for nuclear cardiology procedures Part 1. *J. Nucl. Card.* 1996; vol. 3:G1–G46.
30. Jin M, Yang Y, Niu X, Marin T, Brankov JG, Feng B, Pretorius PH, King MA, Wernick NM. Quantitative evaluation study of four-dimensional gated cardiac SPECT reconstruction. *Phys. Med. Biol.* 2009; vol. 54:5643–5659. [PubMed: 19724094]

Biographies

Ling Li was born in China in 1982. She received the B.S. and M.S. degrees in electrical engineering from the Xi'an Jiaotong University, China, in 2004 and 2006, respectively, and the M.S. degree in electrical and computer engineering from the Illinois Institute of Technology, Chicago, in 2009.

She is currently with the Signal Processing Lab at Ecole Supélec, Paris, France. Her research interests are in medical imaging and signal processing.

Yongyi Yang (M'97–SM'03) received the B.S.E.E. and M.S.E.E. degrees from Northern Jiaotong University, Beijing, China, in 1985 and 1988, respectively, and the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering from the Illinois Institute of Technology (IIT), Chicago, in 1992 and 1994, respectively.

Dr. Yang is currently on the faculty of the Department of Electrical and Computer Engineering, IIT, where he is a Professor. Prior to this position, he was a faculty member with the Institute of Information Science, Northern Jiaotong University, China. His research interests are in signal and image processing, medical imaging, machine learning, pattern recognition, and biomedical applications. He is a coauthor of *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics* (Wiley, 1998).

Dr. Yang is an Associate Editor of the IEEE Transactions on Image Processing.

Fig. 1.

Image sequence with translational motion: (a) noiseless images; (b) noisy images with $\text{SNR} = 7.72$ dB. Frames #1, #3, #5, #7, and #9 are shown.

Fig. 2.
Known motion fields for two example frames (#3).

Fig. 3. Motion-compensated prediction error E_{MC} obtained with different values of model order L and smoothing parameter (Experiment 1).

Fig. 4. Estimated motion fields by classical OFE and the proposed approach (OFE-DFT) for frames #3 at different noise levels. (a) $\text{SNR} = 7.72 \text{ dB}$; (b) $\text{SNR} = 1.63 \text{ dB}$. The OFE is shown in the first row and OFE-DFT is in the second row in both (a) and (b).

Fig. 5. Image sequence with convergent/divergent motion fields: (a) noiseless images; (b) noisy images with $\text{SNR} = 1.05$ dB. Frames #1, #3, #5, #7, and #9 are shown.

Fig. 6.
Known motion fields for two example frames (#3).

Fig. 7.

Motion-compensated prediction error E_{MC} obtained with different values of model order L and smoothing parameter (Experiment 2).

Fig. 8. Estimated motion fields by classical OFE and the proposed approach (OFE-DFT) for frames #3 at different noise levels. (a) SNR = 6.18 dB; (b) SNR = 1.05 dB. The OFE is shown in the first row and OFE-DFT is in the second row in both (a) and (b).

Fig. 9.

Six selected frames (#1, #3, #7, #10, #12, #15) of the myocardium (short axis view) in the cardiac cycle: (a) noiseless reference images; (b) noisy reconstruction images; and (c) motion-compensated filtering of the noisy images using motion from OFE-DFT. The average SNR of 16 frames is improved from 10.06 to 14.56 dB.

Fig. 10.

Motion-compensated prediction error E_{MC} obtained with different values of model order L and smoothing parameter (Experiment 3).

Fig. 11.

Estimated motion fields for four selected gate frames: *First row*: reference motion; *second row*: OFE method; *third row*: proposed OFE-DFT method.

TABLE I

Motion-Compensated Prediction Error E_{MC} (in Decibels) Obtained With Different Model Order L at Different SNR Levels. For Comparison, Results are Also Given for the Classical OFE

SNR	OFE-DFT			OFE
	$L=2$	$L=3$	$L=4$	
13.68 dB	29.64	30.08	29.09	27.66
7.72 dB	28.51	29.52	28.38	26.37
1.63 dB	26.49	27.68	26.42	23.94

TABLE II

Motion-Compensated Prediction Error E_{MC} (in dB) Obtained With Different Model Order L for Different SNR Levels. Results are Also Given for the Classical OFE for Comparison

SNR	OFE-DFT			OFE
	$L = 2$	$L = 3$	$L = 4$	
12.13 dB	27.25	27.64	27.20	26.26
6.18 dB	25.74	26.30	26.03	23.07
1.05 dB	25.70	24.69	23.61	22.93