Commutability of Blur and Affine Warping in Super-Resolution With Application to Joint Estimation of Triple-Coupled Variables

Xuesong Zhang, Jing Jiang, and Silong Peng

Abstract—This paper proposes a new approach to the image blind super-resolution (BSR) problem in the case of affine interframe motion. Although the tasks of image registration, blur identification, and high-resolution (HR) image reconstruction are coupled in the imaging process, when dealing with nonisometric interframe motion or without the exact knowledge of the blurring process, classic SR techniques have to tackle them (maybe in some combinations) separately. The main difficulty is that state-of-the-art deconvolution methods cannot be straightforwardly generalized to cope with the space-variant motion. We prove that the operators of affine warping and blur commute with some additional transforms and derive an equivalent form of the BSR observation model. Using this equivalent form, we develop an iterative algorithm to jointly estimate the triple-coupled variables, i.e., the motion parameters, blur kernels, and HR image. Experiments on synthetic and real-life images illustrate the performance of the proposed technique in modeling the space-variant degradation process and restoring local textures.

Index Terms—Affine motion, blind super-resolution (BSR), blur identification, motion estimation, regularization.

I. INTRODUCTION

Due to the limited resolution of imaging devices, some critical information might be degraded or missed completely in the image acquisition process, such as in the areas of remote sensing, video surveillance, medical imaging, and military target detection and recognition, to name only a few. Image super-resolution (SR) aims at reconstructing a deblurred image with higher optical resolution from multiple blurry low-resolution (LR) observations of the original scene. The subject of SR has been extensively studied in the past three decades and still has been a very active field for its evident practical importance and theoretical interest. Numerous methods have been proposed in the literature (see [1]–[3] for a broad review of the history of SR, related techniques, and recent advances in this area).

Two major factors causing image degradation, as well as two problematic issues, are blurring and decimation, which result in the ill-posedness of the image restoration problem and the paucity of data, respectively. Therefore, it will be an extremely ill-posed inverse problem if one attempts to reconstruct an HR image from a single downsampled version of it. Most existing approaches to SR utilize the multichannel formulation to explore the complementary information among different LR images of the scene, which introduces the third problem in SR, i.e., image registration, to fuse the multiple LR images scattered at irregular positions in the image plane. Thus, a useful SR technique in practice should simultaneously fulfill three tasks, namely, image registration, blur identification, and image reconstruction, which are the objective of our work.

The aforementioned tasks are highly coupled in the image formation process, which imposes a big challenge on coping with the SR problem with in-depth analysis. To circumvent this difficulty, most previous SR algorithms take various measures to reduce the complexity of the problem. Some earlier methods directly decouple the triple-coupled problem into three independent subproblems, i.e., registration, interpolation, and restoration, for example [4]–[6]. Since each subproblem is independently tackled in one stage, more complex motion models can be adopted, such as affine transform [7], projective transform [8], or even optical flow [9]. However, latter research studies in [10]–[12] have shown that subpixel accuracy can be hardly obtained through the registration of aliased LR images, and the resulted catastrophic registration errors will make the SR reconstruction break down. In view of this point, recently, another class of SR approaches has combined image registration and fusion (as well as deblurring) into a common estimation framework and has jointly searched for these estimates, for instance, [10], [12]–[16], and [42], in which the blurring process is assumed known a priori. Obviously, this assumption cannot be satisfied under many practical circumstances and thus limits the efficacy of this kind of algorithms. Typical SR algorithms require image registration with subpixel accuracy, which is difficult to achieve when complex motions are contained in image sequences. A new class of SR algorithms has been proposed to achieve SR on arbitrary sequences while avoiding explicit (subpixel-accurate) motion estimation by means of the nonlocal
means algorithm [43], space–time steering kernel regression [44], and probabilistic local motion estimation [45] for rough motion compensation.

The blur identification methods in published works on blind SR (BSR) can be divided into two categories, namely, parametric blur estimation and nonparametric blur estimation. Nguyen et al. [17] adopted the Lanczos algorithm and Gauss quadrature theory to efficiently compute the large-scale quadratic forms involved in estimating the Gaussian point-spread function (PSF) parameters and the regularization parameters derived by the generalized cross-validation method. The similar approach to the large-scale matrix computation has been also applied to the parametric blur estimation in [18] using the variable projection method [19]. Both of the two BSR techniques assume that the registration parameters are known, which implies a separated image registration stage. He et al. [20] extended the single Gaussian blur type to a learning set composed of different parametric blur models and iteratively estimated the HR image and blur functions under a maximum a posteriori framework. Another sort of parametric blur estimation methods is the newly emerging learning-based blur identification [21]–[23]. The rationale of learning-based methods is to patch wisely super-resolve the HR image via searching in the training database for the most “probable” HR patches, which after degradation resemble the input LR patches the most under some metric. Nevertheless, the extension of blur types cannot entirely break through the limitation of parametric blur estimation as it is impossible to completely cover all blurring cases using only several blur types. BSR with nonparametric blur identification is more difficult mainly because the classic image restoration methods [24]–[26], due to decimation, cannot be applied to BSR in a parallel manner. In the literature, a common assumption of nonparametric BSR is that only pure translation motion exists between consecutive LR frames [27]–[30]. The benefit of this assumption is that translation warping can be represented by a 2-D convolution operation, and thus, the registration job can be incorporated into the blur estimation process. Taking this advantage, Sroubek et al. [29] introduced the 2-D EVAM algorithm for multichannel blind deconvolution (MBD) problem [26] into the BSR framework and utilized the null space of blurs as its regularization term to guarantee consistency of the solution. Assuming the sensor’s PSF and the subpixel accurate shifts between images have been known, Sroubek et al. [46] have recently proposed a fast total variation (TV) regularized SR algorithm that is implemented using fast Fourier transform.

In this paper, we consider a more complicated scenario in which affine transform is used to describe the interframe motion and more complex geometric warps can be approximated by the piecewise affine transforms, such as the methodology in active appearance model [31], [32]. To our knowledge, this is the first work on BSR that simultaneously addresses the nonsymmetric motion estimation, nonparametric blur identification, and HR image reconstruction under a unified framework. The main contribution of this paper is twofold: First, we have proven an equivalent form of the BSR observation model, which leads to the commutability of the space-invariant blurring and affine warping operators, as well as the treatment of decimation as an image scaling down. This conclusion allows for the generalization of the 2-D EVAM-based blur estimation method in [26] and [29] to the affine motion case discussed here. Second, applying this commutability to the BSR problem, we have developed an iterative algorithm to jointly estimate the motion vectors, blur kernels, and HR image. The proposed algorithm adopts a two-layer optimization strategy, which, in the first layer, reduces the triple-coupled BSR problem to a quadratic form with respect to blurs and a nonlinear least-squares (NLS) problem of the motion and HR image and then solves the NLS problem using a Gauss–Newton-based method in the second layer.

The rest of this paper is organized as follows. The problem formulation of image BSR is introduced in Section II. Section III presents the equivalent form of the observation model in both continuous and discrete cases. This serves the blur identification in Section IV. We propose a joint estimation algorithm in Section V. Experimental results are presented in Section VI, and Section VII concludes this paper.

II. PROBLEM FORMULATION

A. Notations and Definitions

We discuss the BSR problem on image plane \( \mathcal{P} \) with Cartesian coordinates \((x, y)\). Bold italic lowercase characters (respectively, uppercase characters) denote column vectors (respectively, matrices). The scene of interest is represented by irradiance field \( f(x) \) injected on \( \mathcal{P} \), where \( x = [x, y]^T \in \mathbb{R}^2 \) denotes the continuous coordinates, whereas \( i = [i_x, i_y]^T \in \mathbb{Z}^2 \) and \( p = [p_1, p_2]^T \in \mathbb{Z}^2 \) denote the discrete positions on HR and LR image grids \( \mathcal{G}_\Delta \) and \( \mathcal{G}_{\Delta'} \), respectively, with \( \Delta' \) and \( \Delta \) being the sampling intervals. \( f(x) \) after discretization yields a 2-D sequence \( f[i] \) arranged into a corresponding matrix \( F \) of size \( m_f \times n_f \). Vector \( f \) is obtained by stacking up the columns of \( F \) lexicographically, i.e., \( f = \text{vec}(F) \) and \( F = \text{Unvec}_{m_f,n_f}(f) \), following the definition in [33].

When we limit the blurring process to be space invariant, the degradation can be modeled as a 2-D convolution of \( F \) with a rectangular blur \( H \) of size \( m_h \times n_h \), i.e., blurred image \( Z = F \ast H \). The matrix–vector form can be written as

\[
zh = \mathbf{C}_{m_f,n_f}(H)f
\]

where \( z \triangleq \text{vec}(Z) \), and 2-D convolution matrix \( \mathbf{C}_{m_f,n_f}(H) \) denotes the convolution of \( H \) with an image \( f \) of size \( m_f \times n_f \). \( \mathbf{C}_{m_f,n_f}(H) \), which was constructed from the columns of \( H \), is a Toeplitz–block–Toeplitz (TBT) matrix of size \( \{m_f + m_h - 1\} \times \{n_f + n_h - 1\} \times m_f \times n_f \). For the specific construction methodology of convolution matrices, the readers are referred to [33].

We define warping operator \( W(.) \) for image \( f(x) \) based on the brightness constancy assumption [34] as follows:

\[
W(f)(x) \triangleq f \left( W^{-1}(x) \right)
\]

where \( W(x) \) represents the geometric transform of pixel’s coordinates, and \( W^{-1}(x) \) is its inverse. Since affine motion is under consideration in this paper, the corresponding affine transform is

\[
W(x) = Mx + d
\]
where $\mathbf{M}$ is the $2 \times 2$ linear transform matrix, and $\mathbf{d}$ is the $2 \times 1$ translation vector. Viewing the elements of $\mathbf{M}$ and $\mathbf{d}$ as the parameters of $\mathbf{W}(\cdot)$, we define the inverse transform as follows:

$$
\mathbf{W}^{-1}(\mathbf{x}; \theta) \triangleq \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \begin{bmatrix} 1 + \theta_1 & \theta_3 \\ \theta_2 & 1 + \theta_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (3)
$$

where $\theta = [\theta_1, \ldots, \theta_6]^T$ is the parameter vector of $\mathbf{W}^{-1}(\cdot)$. Throughout this paper, we will use (2) or (3) in the corresponding context, with $\mathbf{M}$ and $\mathbf{d}$.

### B. Observation Model

Although warping, blurring, and downsampling are the three factors widely accepted by the SR community, in the literature, there exist some ambiguities about the order of warping and blurring in the observation model, e.g., the discussions in [35]–[37]. As most research did, we adopt the warping–blurring–downsampling order in our model to relate HR $f(x)$ and its multiple observed LR versions’ because this formulation more physically models the acquisition process. To get the mathematical convenience without significant deviation from the nature of problem interested, we model the whole blurring effects, including detector’s integration, PSF of lens, motion in aperture time, atmospheric turbulence, etc., to be a single space-invariant blur $h(x)$, i.e., convolution of $f(x)$ with $h(x)$. Thus, the observation model is written as

$$
\mathbf{g}[\mathbf{p}] = S \downarrow ([h \ast W(f)])(x) + n(\mathbf{p}) \quad (4)
$$

where $W(\cdot)$ is the image warping operator defined in (1) and (3) to accommodate the relative camera-scene motion such as scaling, rotation, translation, and shear, $S \downarrow$ denotes the downsampling operator, and $n(\mathbf{p})$ is additive noise.

Given a sequence of blurred and noisy LR images $g_k[\mathbf{p}], 1 \leq k \leq K$, we address the task of restoring an HR representation of original scene $f(x)$ with the simultaneous estimation of blurs $\hat{h}_k$ and affine motion parameters $\theta_k$.

### C. Related Work

We first briefly review two blur identification methods that are relevant to our work. Omitting the downsampling and warping operators in (4), we will face an MBD problem that recovers original image $\mathbf{F}$ and blur kernels $\mathbf{H}_k$’s given $K$ measurements $\mathbf{G}_k$’s, i.e.,

$$
\mathbf{G}_k = \mathbf{H}_k \ast \mathbf{F} + \mathbf{N}_k, \quad k = 1, \ldots, K \quad (5)
$$

where $\mathbf{N}_k$ is the noise matrix. Assembling the following equations in the noiseless case

$$
\mathbf{H}_k \ast \mathbf{G}_l - \mathbf{H}_l \ast \mathbf{G}_k = 0, \quad 1 \leq k < l \leq K \quad (6)
$$

into a linear system, Harikumar and Bresler [33] provided a state-of-the-art MBD solution, i.e.,

$$
\mathbf{N}_{\text{MBD}} \mathbf{h} = 0 \quad (7)
$$

where $\mathbf{N}_{\text{MBD}}$ is the $K \times (K+1)$ matrix, $\mathbf{h}$ is the parameter vector of $\mathbf{W}(\cdot)$, $\mathbf{Q}_k$ is the $K \times K$ matrix for $k = 1, \ldots, K - 1$, and $\mathbf{G}_k$ is the $K \times K$ matrix for $k = 1, \ldots, K - 1$.

What hinders us from directly estimating the blurs using (7) for the BSR case is the existence of image warping operator $W(\cdot)$ and downsampling operator $S \downarrow$ in (4) because they violate the commutability required by (6). Limiting the motion model to the pure translation case (thus warping commutes with convolution), Sroubek et al. [29] tackled the downsampling problem by upsampling the null space of the “LR” version of blurs, i.e.,

$$
\tilde{\eta}_{kn} = \mathbf{D}^T \eta_{kn} \quad (8)
$$

where $\mathbf{D}$ is the 2-D decimation matrix with the downsampling factor being $s$, and $\eta_{kn}$, for $k = 1, \ldots, K$ and $n = 1, \ldots, n - \text{nulity}(\mathbf{G})$, are the nullifying filters of size $(m_s, n_s)$, which are made of values of the null space of $\mathbf{G} \triangleq \{C_{(m_s, n_s)}[\mathbf{G}_1], \ldots, C_{(m_s, n_s)}[\mathbf{G}_K]\}$. The following same form equation as (7) can then be obtained:

$$
\mathbf{N}_{\text{BSR}} \mathbf{h} = 0 \quad (9)
$$

where

$$
\mathbf{N}_{\text{BSR}} \triangleq \begin{bmatrix}
C_{(m_s, n_s)}[\tilde{\eta}_{11}] & \cdots & C_{(m_s, n_s)}[\tilde{\eta}_{1K}]
\vdots & \ddots & \vdots
C_{(m_s, n_s)}[\tilde{\eta}_{K1}] & \cdots & C_{(m_s, n_s)}[\tilde{\eta}_{KK}]
\end{bmatrix}.
$$

In Section III, we will prove an equivalent form of the BSR observation model, which leads to 1) the commutability of the convolution and affine warping operators, and 2) the treatment of decimation as an image scaling down. These properties will facilitate the necessary modifications of (7) and (9) for the case of affine motion in Section IV.

### III. EQUIVALENT OBSERVATION MODEL

Here, we explore the feasibility and conditions to change the order of warping and blur operators in order to simplify the BSR problem defined in the previous section. We start from the establishment of an equivalent form of (4) under continuous coordinates in the absence of noise. Then, this result is extended to the discrete case.

#### A. Continuous Model

The following theorem proves that affine warping an image then blurring it is equivalent to inversely warping the blur first then convolving with the original image and finally warping the intermediate blurred image in addition to a scaling of intensity.

**Theorem 1:** Let $\mathcal{D}_\Delta(z(x)); i \in \mathbb{Z}^2$, be the decimation operator, with $\Delta$ denoting the sampling interval, and $W_s(z(x)) = z(W_s^{-1}(x))$ be the image scaling operator, with $s \in \mathbb{R}^+$ denoting the scaling factor, where $z(x)$ is a continuous image and the scaling transform $W_s(x) = \text{diag}[1/s, 1/s]x$. 
Then, in the noiseless case, observation model (4) has the equivalent
\[ g[p] = D_{\Delta s} \left( \frac{1}{\det(M)} W_s \left( \left[ f * W^{-1}(h') \right] \right)(x) \right). \] (10)

**Proof:** See Appendix A.

Corollary 1: Define \[ g(x) \triangleq \frac{1}{\det(M)} W_s(W([f * W^{-1}(h')])(x) \]. Then, the following equation holds:
\[ W^{-1}(W^{-1}(g))(x) = [f * W^{-1}(h')] \](x). \] (11)

**Proof:** The establishment of (11) is given by the following inversability of the warping operator:
\[ W^{-1}(W(f))(x) = W(f) \left( W^{-1}(x) \right) = f(x). \]

Corollary 1 and its discrete version (refer to (14) in the following part) will help us construct a set of blurred HR images for the exploration of the solution space of blurs, which will be used in Section IV.

### B. Discrete Model

When transforming the continuous model to its discrete version, one has to analyze the sampling condition and the sampling interval that should be chosen to avoid aliasing as much as possible. From the detailed discussion in Appendix B, we conclude that aliasing free is guaranteed when sampling \( W(f)(x) \) with interval \( \Delta' \) if
\[ \Delta'/\sqrt{\det(M)} \leq \Delta_{\text{Nyquist}} \] (12)
where \( \Delta_{\text{Nyquist}} \) denotes the Nyquist sampling interval.

Accordingly, in the noiseless case, (4) can be written as
\[ g = DHWF \] (13)
where \( D \) is the 2-D down-sampling matrix with factor \( s \in N^+, \) and \( H \triangleq C_{(m,r,n)} \{ H \} \) is the convolution matrix constructed from \( H \).

Nevertheless, the inevitable aliasing occurs in down-sampling process \( g[p] = D_{\Delta s}([h * W'(f)](x)) \) in (4) or, equivalently, in (10) when taking \( \Delta = \Delta' \) as the LR sampling interval instead of \( \Delta' \), because \( \Delta = \Delta'/\sqrt{\det(M)} \gg \Delta_{\text{Nyquist}} \) violates the Nyquist theorem. Therefore, uniformly sampling \( g[x] \) defined in Corollary 1 with \( \Delta \) will incur aliasing. Due to the blurring effect of \( h(x) \) to \( W(f)(x) \), however, a part of high-frequency components in \( W(f)(x) \) get lost. Therefore, when \( s \) is relatively small, it is acceptable to reconstruct \( g(x) \) approximately from \( g[p] \), i.e.,
\[ g(x) \approx \sum_{p \in \Delta} g[p] \varphi(x - p\Delta), \] where \( \varphi(x) \) is an interpolation kernel whose shifted duplications generate a basis on \( \mathbb{R}^2 \). Thus, the matrix–vector form of (11) is written as
\[ \frac{1}{\det(M)} W_s^{-1} W_s^{-1} g = F W^{-1} h \] (14)
where \( W_s^{-1}, W^{-1} \), and \( W_s^{-1} \) are the corresponding interpolation matrices of \( W_s^{-1}(g)(x), W^{-1}(h)(x) \), and \( W^{-1}(W_s^{-1}(g))(x) \), respectively; \( F \triangleq C_{(s,n)} \{ F \} \) is the convolution matrix constructed from \( F \). Notice that \( W_s^{-1} \) is an interpolation or decimation matrix depending on whether \( s > 1 \) or \( s < 1 \) (therefore not square and \( W_s^{-1} \) is not the identity matrix because of the aforementioned aliasing) and that \( W_s^{-1} \) and \( W^{-1} \) are the ordinary square image warping matrices. Of course, \( W_s^{-1} \) and \( W^{-1} \) can be combined into a single matrix, but we retain both of them here to emphasize the two-step interpolation procedures of the LR images: first magnifying \( g \) times then inversely warping the magnified image \( g \).

### IV. Blur Identification

Consider \( K \) blur filters \( h_k \) and \( h_k = \text{vec} \{ H_k \} \), \( 1 \leq k \leq K \). We assume the size of the filters to be \( m_k \times n_k \). Here, we first explore subspace \( \mathcal{S} \) orthogonal to \( h \) and \( h ' \) defined in (14) and (15), which will be utilized as a regularization constraint on \( h \). Then, we address the task of efficiently solving for the eigenvectors of \( \mathcal{S} \) and finally present the estimation method of the blur size.

#### A. Subspace-Based Constraint on Blurs

Consider (14), we define a blurred HR image \( y_k \) through the interpolation and registration of LR image \( g \) to HR grid \( \mathcal{G} \), i.e.,
\[ y_k = \frac{1}{\det(M_k)} W_{s_k}^{-1} W_{s_k}^{-1} g_k. \] (15)
Substituting \( y_k \) into (14), we have
\[ y_k = F W_{s_k}^{-1} h_k. \] (16)

Inspired by the subspace techniques developed in [29] and [33], we explore the orthogonal complement of the space of blurs to obtain their optimal estimates in the sense of least-squares (LS) error. Introducing an intermediate kernel \( E \) of size \( m_r \times n_r \) and defining
\[ \mathcal{Y} \triangleq \left[ C_{(m,r,n)} \{ Y_1 \}, \ldots, C_{(m,r,n)} \{ Y_K \} \right] \]
\[ \mathcal{H} ' \triangleq \left[ C_{(m,r,n)} \{ W_{1}^{-1} h_1 \}, \ldots, C_{(m,r,n)} \{ W_{K}^{-1} h_K \} \right] \] (17)
we can get \( \text{Null}(\mathcal{Y}) = \text{Null}(\mathcal{H} ') \), which is similar to the results in [29]. Then, the null-space matrix for the case of affine motion corresponding to (9) can be constructed as (18), shown at the bottom of the page, where \( \eta_{1,k} \triangleq \eta_{1,1}, \ldots, \eta_{1,K} \) for \( r = 1, \ldots, R \) and \( k = 1, \ldots, K \), and \( \eta_{1,k} \)’s are the eigenvectors of \( \text{Null}(\mathcal{Y}) \) with \( R \triangleq \text{nullity}(\mathcal{Y}) = K m_r n_r - (m_h + m_r - \ldots) \)
1)[n_h + n_e - 1]. The size of $\mathcal{N}_\text{BSR}$ is $R(m_k + m_e - 1)(n_h + n_e - 1) \times K m_n n_k$. To reduce the uncertainty of $h$, we try to reduce nullity of $\mathcal{N}_\text{BSR}^\dagger$, which leads to the following constraints on $(m_e, n_e)$:

\[
\begin{align*}
&\{K m_n n_e > (m_k + m_e - 1)(n_h + n_e - 1) \\
&\{R(m_h + n_e - 1)(n_h + n_e - 1) \approx K m_n n_h.
\end{align*}
\]

Defining $h'_k = W_{k}^{-1} h_k$ in (16), the null-space matrix $\mathcal{N}_\text{BSR}^\dagger$ corresponding to (7) can be constructed by substituting $C_{\{m_k, n_k\}}\{G_k\}$ with $C_{\{m_n, n\}}\{Y_k\}$. In this scenario, we in fact estimate warped blurs $h'_k$'s instead of the original ones, which can be obtained by $h_k = W_k h'_k$.

For the coherent expression of our joint estimation of $h$, $f$, and $\theta$, we take $\mathcal{N} = \mathcal{N}_\text{BSR}$ in the rest of this paper and get a system of linear equations, i.e.,

\[
\mathcal{N} h = 0.
\] (19)

Due to noise and interpolation errors, we adopt the LS technique for the blur estimation problem as follows:

\[
\hat{h} = \arg \min_h |\mathcal{N} h|^2.
\] (20)

### B. Implementation Issues

For blur kernels with ordinary size (for example, $10 \times 10$), $\mathcal{N}^\dagger \mathcal{N}$ in (20) will be of moderate scale ($\sim 10^3 \times 10^3$). Thus, (20) can be solved by singular value decomposition (SVD) without heavy computational burden, whereas $Y$ is too large, size $(m_g + m_e - 1)(n_g + n_e - 1) \times K m_n n_e$, to directly compute its null space Null($\mathcal{Y}$). We circumvent this difficulty by performing the SVD of $\mathcal{Y}$ since $\mathcal{Y}^T \mathcal{Y}$ is much smaller. Eigenvectors corresponding to the $R$ smallest eigenvalues of $\mathcal{Y}^T \mathcal{Y}$ span Null($\mathcal{Y}$). By means of the efficient calculation skills of the matrix product in [33], the computational complexity reduces from $O(K^2 m_g^2 n_g^2 m_{\nu} n_{\nu})$ to $O(K^2 m_n n_e \log(m_{\nu} n_{\nu}))$.

Another issue we have to address is the estimation of blur size $(m_k, n_k)$ before we estimate the blur filters themselves. We tackle this by converting the BSR problem into an MBD problem. By (16), we have

\[
Y_k = f * H'_k, \quad 1 \leq k \leq K
\] (21)

where $Y_k = \text{Unvec}_{m_{\nu}} \{y_k\}$, and $H'_k = \text{Unvec}_{m_{\nu}} \{W_k^{-1} h_k\}$. This is an MBD problem that recovers original image $F$ and warped blurs $H_k$ of the filter size, the output of the blur estimation will be $U_k = H_k(M_k, N_k) \ast B$, where $B$ is a spurious common factor of size $(M_k - m_k + 1, N_k - n_k + 1)$; on the other hand, the underestimate of blur sizes will make the problem unsolvable [33]. Consequently, we estimate $(m_k, n_k)$ by minimizing the following residual function $\psi$ with respect to $(M_k, N_k)$, which is sensitive to overestimation of blur sizes if the noise level is not too high (see the proof in [33]):

\[
(m_k, n_k) = \arg \min_{(M_k, N_k)} \psi(M_k, N_k)
\] (22)

where

\[
\psi(M_k, N_k) = \min_{Z} \sum_{k=1}^{K} \left| Y_k - Z * H'_k(M_k, N_k) \right|^2.
\] (23)

### V. Joint Estimation Algorithm

Here, we propose a joint estimation method of affine motion parameters $\theta$, blur kernels $h$, and HR image $f$. Because the blurring effect in the image formation process can be modeled as convolution of the image with a compact support blur kernel (a Fredholm integral equation of the first kind [38]), the BSR problem is mathematically ill-posed in terms of both $f$ and $h$. We adopt regularization techniques in our method to obtain numerically stable and meaningful results. Specifically, we estimate $(f, \theta, h)$ by minimizing the following energy function:

\[
\Phi(f, \theta, h) = \sum_{k=1}^{K} \left| \mathcal{D}h_k W(\theta_k) f - g_k \right|^2 + \lambda_1 \left| L f \right|^2 + \lambda_2 \left| L h \right|^2 + \lambda_3 \left| L \mathcal{N} h \right|^2
\]

where $B \Delta = \left[ (\mathcal{D}h_1 W(\theta_1) f)^T, \ldots, (\mathcal{D}h_K W(\theta_K) f)^T \right]^T$, $g \Delta = [g_1^T, \ldots, g_K^T]^T$, $L$ denotes the Laplacian filter, and $\lambda_{1-3}$ are the regularization parameters. Throughout this paper, we emphasize the space-variant warping matrix by $W(\theta_k)$ with motion parameters $\theta_k$, and $W^{-1}(\theta_k)$ represents the inverse warping matrix of appropriate size.

### A. Hierarchical Optimization Strategy

The main difficulty of the minimization of (24) resides in the fact that $f$, $\theta$, and $h$ are triple coupled due to the warping and convolution operations, hence leading to the nonconvexity of $\Phi(f, \theta, h)$. We solve this problem in a tractable way by means of a two-layer hierarchical optimization strategy. By inspection of the first term (data fidelity) of (24), $h$ can be decoupled from $(f, \theta)$ by exchanging the positions of $h$ and $f$ (following Theorem 1), i.e.,

\[
\Phi(f, \theta, h) = \sum_{k=1}^{K} \|\text{det}(M_k) \mathcal{D}W(\theta_k) F W^{-1}(\theta_k) h_k - g_k \|^2 + \lambda_1 \| L f \|^2 + \lambda_2 \| L h \|^2 + \lambda_3 \| L \mathcal{N} h \|^2.
\] (25)

Notice that $W(\theta_k)$ and $W^{-1}(\theta_k)$ in (25) are of different sizes because they represent the warping of a blurred HR image $f \ast W(h_k)$ and the inverse warping of a blur kernel $h_k$ under parameters $\theta_k$, respectively. In the rest of this paper, we will choose (24) or (25) for convenience to derive $(f, \theta, h)$. Taking advantage of this relation, we adopt the alternating minimization method to decompose the original tough problem into two relatively simpler procedures, i.e., one quadratic form w.r.t. $h$

\[
\Phi_1(h; f^{(t)}, \theta^{(t)})
\]

\[
= \sum_{k=1}^{K} \| \text{det}(M_k^{(t)}) \mathcal{D}W(\theta_k^{(t)}) \mathcal{Z}^{(t)} W^{-1}(\theta_k^{(t)}) h_k - g_k \|^2 + \lambda_2 \| L h \|^2 + \lambda_3 \| L \mathcal{N} h \|^2
\] (26)
TABLE I

TWO-LAYER OPTIMIZATION STRATEGY

Step 1. Estimating \( h^{(0)} \) using the current estimates \( (f^{(0)}, \theta^{(0)}) \):

1. Align and interpolate \( \{g_k\}_{k=1}^K \) to obtain \( \{y_k\}_{k=1}^K \) using (15).
2. Compute \( \text{Null}(\mathbf{Y}) \) following the method in Section IV where \( \mathbf{Y} \) is defined in (17), and construct \( \mathbf{N} \) by (18).
3. Solve \( h^{(0)} = \arg \min_{h} \Phi(f^{(0)}, \theta^{(0)}, h) \) by setting \( \frac{\partial \Phi}{\partial h} = 0 \) and adding
   normalization constraint to \( h \), which is equivalent to solving
   \[
   (A^T A + \lambda_2 A^T N^T + \lambda_3 L^T L) h = A^T g
   \]
   where
   \[
   A = \text{bldkdg} \left[ \left[ \text{det} (M_k^T) \text{bw} (\theta_k^T) \text{e}^{-i \phi (y_k)} \right]_{k=1}^K \right],
   \]
   and \( g = \left[ g_1^T, \ldots, g_K^T \right]^T \).
4. Compute the residual function w.r.t. \( (M_k, N_k) \) defined as follows:
   \[
   \nu (M_k, N_k) = \sum_{k=1}^K \text{bw} (\theta_k^T) H^{(0)} (\theta_k - f^{(0)} - g_k)^T.
   \]
5. Reduce \( (M_k, N_k) \) by 1, and repeat (1) to (4). Determine the true filter size \( (m, n) \) by (22) and take \( h^{(0)} = h^{(m,n)} \).

Step 2. Estimating \( (f^{(0)}, \theta^{(0)}) \) using \( (f^{(0)}, \theta^{(0)}, h^{(0)}) \):

1. By the Gauss–Newton algorithm, \( \min_{(f, \theta, h^{(0)})} \Phi(f, \theta, h^{(0)}) \) is transformed into a linear system \( Mz = \beta \). We present the notations and derivations of this equation in Appendix C.
2. Solve \( Mz = \beta \) by preconditioned conjugate gradient (PCG) method and update \( (f^{(0)}, \theta^{(0)}) \) by:
   \[
   f^{(0)} = f^{(0)} + \beta f, \theta^{(0)} = \theta^{(0)} + \beta \theta .
   \]
when \( (f, \theta) \) are fixed, of which the closed-form solution is straightforward (see Table I), and one NLS problem of \( (f, \theta) \)
\[
\Phi_2 (f, \theta; h^{(1)}) = \sum_{k=1}^K \left| D H_k^T (W_k (\theta_k)) f - g_k \right|^2 + \lambda_1 \left| Lf \right|^2 .
\]
when \( h \) is fixed. In the second layer, since \( \theta \) and \( f \) are still coupled in (27) by the warping operation, we estimate them simultaneously by the Gauss–Newton algorithm (see Appendix C).

The proposed algorithm alternates between steps 1 and 2 in Table I until some convergence criterion or a maximum number of iterations has been reached. Although \( \Phi(f, \theta, h) \) is not convex in the three dimensions, and thus the convergence to the global minimum is not guaranteed, our extensive testing shows that the decomposition of \( \min_{(f, \theta, h)} \Phi(f, \theta, h) \) into \( \min_{h} \Phi_1 (h; f^{(0)}, \theta^{(0)}) \) and \( \min_{f, \theta} \Phi_2 (f, \theta; h^{(1)}) \) can yield relatively good results. To alleviate the computational burden, substeps 4 and 5 in step 1 could be conducted only in the first loop and keep the filter size estimate throughout the whole optimization process. Many tests indicate that this simplification does not introduce serious errors of the filter sizes.

B. Choice of Regularization Parameters

\( \lambda_1 \sim \lambda_2 \) are automatically determined in subproblems \( \min_{(f, \theta, h)} \Phi_2 (f, \theta, h^{(1)}) \) and \( \min_{h} \Phi_1 (h; f^{(0)}, \theta^{(0)}) \). Among the three regularization parameters, \( \lambda_1 \) and \( \lambda_2 \) are more critical to the SRR result than \( \lambda_3 \). In (27), we adopt the linear regularization functional form from [40] and [41], which results in

\[
\lambda_1 = \frac{\| Bf - g \|^2}{\frac{1}{\gamma} - \| Lf \|^2},
\]
where \( 1/\gamma \) is a quantity that controls the convergence of \( \min_{(f, \theta, h)} \Phi_2 (f, \theta, h^{(1)}) \). The justification of this choice of \( \lambda_1 \) can be found in [40] and [41], and the sufficient condition for convergence is

\[
\frac{1}{\gamma} > \| Bf - g \|^2 + \| Lf \|^2 .
\]
What should be noticed is that (29) still holds true for affine warping of images, although the motion model in [41] is pure translation. The only difference is \( \sigma_{\text{max}} (W_k W_k^T) \neq 1 \) because affine transform is space variant, and thus, \( W_k \) is not normalized, where \( \sigma_{\text{max}} (A) \) denotes the maximum singular value of matrix \( A \). However, this does not affect the conclusion in (29).

An upper bound of \( \| Bf - g \|^2 + \| Lf \|^2 \), 2 \( \| g \|^2 \), was chosen to be the value of \( 1/\gamma \) in [40] and [41], which has been verified to be a too big bound that results in noise magnification in our experiments. To overcome less regularization, we choose another tighter bound, i.e.,

\[
\frac{1}{\gamma} = \| Bf - g \|^2 + \left( \frac{\| Bf - g \|^2 + \| Lf \|^2}{4\| Bf - g \|^2} \right)^2
\]
in which we use the following inequality:

\[
\| Lf \|^2 \leq \frac{\| Bf - g \|^2 + \| Lf \|^2}{4\| Bf - g \|^2}.
\]
\( \lambda_3 \) plays a role that makes the contribution of null-space constraint \( \| Nh \|^2 \) comparable with residual energy \( \| Bf - g \|^2 \). Thus, \( \lambda_3 \) is chosen to be proportional to the ratio of \( \| Bf - g \|^2 \) and \( \| Nh \|^2 \). Thus

\[
\lambda_3 = \alpha \times \frac{\| Bf - g \|^2}{\| Nh \|^2},
\]
where \( \alpha \) is a factor that needs to be manually adjusted. \( \lambda_2 \) is determined in a similar way for controlling the smoothness of the blur kernels, i.e.,

\[
\lambda_2 = \beta \times \frac{\| Bf - g \|^2}{\| Lh \|^2},
\]
where \( \beta \) is also a tradeoff factor. In our experiments, \( \alpha \) and \( \beta \) are set to be \( 10^{-3} \) and \( 10^{-2} \), respectively, which usually give satisfying results.

VI. EXPERIMENTAL RESULTS

Here, we will report the results of experiments on both synthetic and real-life images and compare the performance of
the proposed BSR algorithm with existing methods. We first conducted comparisons with the BSR algorithm devised for pure translation motion (denoted by PT-BSR) in [29] and the variational Bayesian SR algorithm (denoted by VBSR) in [42] via Monte Carlo simulations under affine motion and different levels of noise. To evaluate the performance of our algorithm for real BSR problems, we also demonstrated the reconstruction results under different SR factors from 2 to 4, including cases of both compressed and uncompressed sequences.

A. Simulated Data

In this experiment, we cropped a 96 × 96 patch from Lena as the original HR image. To generate the test sequences of LR images, the HR image was warped with different random affine motion vectors, blurred by five randomly generated 8 × 8 kernels (PSFs) whose marginal 2-pixel space was reserved for accommodating warping of the PSFs, and finally downsampled by a factor of \( s = 2 \). The original HR image and simulated PSFs are shown in Fig. 1. To test the robustness of our algorithm to noise, we added to the LR images white Gaussian noise with different signal-to-noise ratios (SNRs) from 50 to 10 dB. We first compare our algorithm (called AT-BSR hereafter) with PT-BSR to see the influence of space-variant warping on blur estimation and image reconstruction. To verify the contribution of blur identification to SR reconstruction, we will use the true PSFs and a 6 × 6 uniform PSF for VBSR to compare their outputs (denoted by VBSR-True and VBSR-Uniform, respectively) with the above BSR algorithms. The peak SNR (PSNR) is used as a quantitative measure in terms of the reconstruction quality, which is defined as

\[
\text{PSNR}(\hat{f}) = 10 \log_{10}(255^2 / \| f - \hat{f} \|^2 / m \times n),
\]

where \( f \) is the estimate of original HR image \( f \). At each SNR noise level, we repeated the simulation ten times, and each time, the generated LR sequence was fed into AT-BSR and PT-BSR in turn. Since the PT-BSR algorithm requires the LR sequence to be roughly registered to leave only small residual translations, we restricted the mean motion vector within two HR pixels for a fair play. The LR sequences for VBSR contain motions of only rotation and translation for a similar consideration. The regularization parameters in PT-BSR were manually adjusted in order to obtain the best reconstruction quality. The average PSNR results are plotted in Fig. 2. As expected, VBSR outperforms PT-BSR and AT-BSR when using the true PSFs in each noise setting and exhibits similar performance as the BSR algorithms when assuming a uniform PSF. A significant improvement of VBSR over the two BSR algorithms can be found in the low SNR case (less than 20 dB) due to the TV prior to its used. However, the PSNR criterion does not coincide with the visual perception of the reconstructed images presented in Fig. 3 when the SNR levels are higher than 20 dB. The reconstructed HR images and the cubic interpolations are shown in Fig. 3. The images are of different sizes because of the different image boundary disposals adopted by the algorithms. It is evident that the AT-BSR algorithm has superiority over PT-BSR and VBSR in the restoration of local textures. The two BSR algorithms yield almost identical reconstruction results in the central parts of images, e.g., the nose, whereas the AT-BSR algorithm outperforms its counterpart in regions far from the center, e.g., the hair and pleats of the hat. This is because the variation of motion vectors in outer regions becomes bigger and bigger, which is a space-variant degradation process and cannot be modeled as a convolution operation (although the amplitudes of motion vectors were small), such as in the PT-BSR algorithm. The proposed AT-BSR algorithm, however, is capable of representing and compensating this kind of space-variant degradation very well. The VBSR algorithm provides good noise suppression performance; however, it seems to do better in edge preservation than in texture restoration probably due to its TV-based regularization, which, in fact, introduces double-fold smoothing of image gradients. We infer from the results in Figs. 2 and 3 that one should presume a reasonable PSF (by our experience, correct blur size is more crucial to the final results than accurate estimation of the PSFs’ values) to conduct SR reconstruction rather than attempt to estimate the blur kernels using very noisy images, which will be further verified by the blur estimation results in Fig. 4.

The estimated PSFs are shown in Fig. 4, where we placed the results of the AT-BSR and PT-BSR algorithms in the top and bottom rows, respectively, in Fig. 4(b)–(f). The shapes of
the blur kernels given by AT-BSR closely agree with the true PSFs in Fig. 4(a) at SNR levels from 50 to 20 dB, but the distributions of intensity have small divergence. The obvious errors take place at 10 dB due to the very heavy noise. Under all conditions of SNR, the estimated PSFs of the first frame are the most accurate because this frame was chosen as the reference image and, thus, not affected by the interpolation error, whereas the blur kernels of other frames have to compensate the registration error. This compensation is fulfilled via the interpolation of blur kernels, e.g., the estimated PSFs in Fig. 4(b)–(d) are shifted by one pixel from the center, which results in some differences between the estimates and the ground truth. Since the registration error is inevitable, the compensation capability of blur kernels to registration, however, helps the reconstructed images be less demanding on the registration accuracy. The presented PSFs given by PT-BSR are the results when the reconstructed images were most visually pleasing. It is somewhat surprising that the intensity distributions of the blur kernels estimated by PT-BSR concentrate on the central one to two pixels, but the images in Fig. 3(b) are relatively good, which probably results from the compensation of a globally translated blur to an affine transform. This can be illustrated using an extreme case that the PSF reduces to a delta function, and in this situation, the blur and warping operators commute whatever image warping is performed.

B. Real Data

We now demonstrate the algorithm on real images, including compressed and uncompressed scenarios. The compressed test sequences, namely, Face 1 and Face 2, were downloaded from MDSP (http://www.soe.ucsc.edu/~milanfar/DataSets). In each experiment, we used the first \( K \) LR frames to reconstruct the HR image, where \( K = s^2 + 1 \) and the decimation factor \( s = 2, 3, 4 \). According to MDSP, these frames approximately follow the global translation motion model. Therefore, the AT-BSR and PT-BSR algorithms give similar results in
Fig. 4. Estimated PSFs. (a) Ground-truth PSFs. (b) SNR = 50 dB. (c) SNR = 40 dB. (d) SNR = 30 dB. (e) SNR = 20 dB. (f) SNR = 13 dB. In (b)–(f), the top rows are the results of AT-BSR, whereas the bottom rows are the results of PT-BSR. The PSFs shown by PT-BSR are the convolutions of the estimated volatile PSFs with the Gaussian sensor PSF with std = 0.34.

Figs. 5(d) and (e) and 6(d) and (e), as expected. Due to the deconvolution operations in the BSR algorithms, the deblurring performances of the two BSR algorithms are better than the results provided by MDSP (http://www.soe.ucsc.edu/~milanfar/ExampleImages). Since the data of the sequences were compressed, the valuable complementary information among consecutive frames was badly mangled, and hence, one cannot observe pronounced improvement of reconstruction quality when the magnification factor is beyond 3.

The next experiment illustrates our algorithm on an uncompressed sequence of infrared frames. An aircraft flying in the sky was captured by a forward-looking infrared camera with a 320 × 240 detector array. We extracted the 20 × 15 regions of interest containing the aircraft as the LR inputs and run the BSR algorithms taking the magnification factor $s = 3, 4$. From the bicubic interpolations in Fig. 7(c), one can observe that the interpolations appear more and more blurry as the SR factor increases. On the contrary, the BSR algorithms PT-BSR and AT-BSR, fusing data from multiple LR frames and performing deconvolution, provide more satisfying reconstructions with clear aircraft profiles in Fig. 7(e) and (f), rather than perform worse as the SR factor increases. For comparison purposes, we also implemented the separated registration–interpolation–restoration algorithm [4]. Notice that the reconstruction artifacts due to misregistration become severe as the magnification increases (accordingly the number of images to be aligned),
whereas the joint estimation methods in AT-BSR and PT-BSR do better at accommodating some registration errors. From Figs. 5–7, we deduce that super-resolving compressed imagery beyond three times will be for the most part a fruitless effort and that one should try to do SR before the compression stage in order to obtain significant improvement of image quality.

Because the motion of the aircraft during the acquisition time is basically pure translation, both the BSR algorithms yield similar outputs, as shown in Fig. 7(e) and (f). To verify the efficacy of AT-BSR in dealing with affine motion, we extracted the license plates from two sequences of moving cars, one of which was a compressed traffic surveillance video [see Fig. 8(a), left] and the other was the uncompressed car sequence from MDSP [see Fig. 8(a), right]. Both of the two sequences approximately follow the affine motion model due to global translations and zooming and were used to test the proposed algorithm. A similar comparison scenario to Fig. 7 was applied, and the results are shown in Fig. 8. Obviously, in the case of affine interframe motion, AT-BSR outperforms the PT-BSR and registration–interpolation–restoration techniques, better recovering the letters and numbers in the license plates.

VII. CONCLUSION

We have shown that the order of blurring and warping operators in SR under affine motion can be exchanged through the rigorous proof of an equivalent form of the continuous observation model, and this conclusion has been extended to the discrete case under some mild conditions. Applying this commutability to the BSR problem, we have developed an iterative algorithm to jointly estimate the motion vectors, blur kernels, and HR image with a two-layer optimization strategy. The efficient implementation method and determination of the regularization parameters were presented. Experimental results on synthetic data and real-life image sequences show the validity of the equivalent observation model and the efficiency of the proposed BSR method in accommodating affine warping. The applicable extensions of the proposed BSR method in the scenarios that local motions and space-variant blurs exist in the scene are currently under investigation.

APPENDIX A

PROOF OF THEOREM 1

We begin with stating and proving two lemmas.

Lemma 1: Let \( z(\mathbf{x}) \) be a 2-D continuous signal. Then, the uniform sampling of \( z(\mathbf{x}) \) with interval \( \Delta \) (assuming \( \Delta_x = \Delta_y = \Delta \)) is equivalent to scaling \( z(\mathbf{x}) \) by a factor of \( s \in \mathbb{R}^+ \) first and then sampling it with interval \( \Delta / s \), i.e.,

\[
D_\Delta (z(\mathbf{x})) = D_{\Delta / s}(W_s(z)(\mathbf{x})).
\]  

Proof: By the definition of image warping in (1), we prove (33) straightforwardly, i.e.,

\[
D_{\Delta / s}(W_s(z)(\mathbf{x}))[i] = D_{\Delta / s}(z[(\mathbf{W}_s^{-1}(\mathbf{x}))[i]] = z(s(i\Delta / s)) = D_\Delta (z(\mathbf{x}))[i].
\]

Lemma 2: For the affine image warping operator \( W(\cdot) \) defined in (1) and (2), the following is true:

\[
h * W(f)(\mathbf{x}) = \det(M) W[f * W^{-1}(h)](\mathbf{x})
\]  

Proof: By changing the integral variables by \( \mathbf{v} = \mathbf{W}^{-1}(\mathbf{u}) \), we have

\[
[h * W(f)](\mathbf{x}) = \int h(\mathbf{u}) \cdot f(\mathbf{W}^{-1}(\mathbf{x} - \mathbf{u})) \, d\mathbf{u}
\]  

\[
= \int \int h(\mathbf{W}(\mathbf{v})) \cdot f(\mathbf{W}^{-1}(\mathbf{x}) - \mathbf{v}) \cdot \det(M) \, d\mathbf{v}
\]  

\[
= \det(M) \int h(\mathbf{W}^{-1}(\mathbf{x})),
\]  

\[
= \det(M) W(h(\mathbf{W}^{-1}(h))),
\]

\[
= \det(M) W(f)(\mathbf{x}).
\]
Notice that, in the preceding derivation, we use 1) the Jacobian of \( u \) w.r.t. \( v \) and 2) the intermediate blur function \( h_W(u) \triangleq h(W(u)) = W^{-1}(h(u)) \).

Now, we proceed to the proof of Theorem 1. Using the results in Lemmas 1 and 2, we obtain

\[
g(p) = D_{\Delta_1} [h \ast W(f)](x) = D_{\Delta_1} [W(x)(h \ast W(f))](x) = D_{\Delta_1} \left[ |\text{det}(M)\right| W_x \left(W \left( [f \ast W^{-1}(h)] \right) \right)(x)
\]

which completes the proof of Theorem 1.

**APPENDIX B**

**DISCUSSION ON SAMPLING CONDITIONS**

Consider we uniformly sample original scene \( f(x) \) and blur function \( h(x) \) with interval \( \Delta' \) in image plane \( \mathcal{P} \), i.e., \( f[i] = D_{\Delta'}(f(x)), h[i] = D_{\Delta'}(h(x)) \). If \( \Delta' \leq \Delta_{\text{Nyquist}} \), then \( f(x) \) and \( h(x) \) can be precisely recovered from their discrete samples.

**Theorem 2:** Define a measure function of \( f(x) \) by the area of its support \( D_{\Delta^*} \), \( \mu(f(x)) \triangleq \int_{x \in D_{\Delta^*}} dx \). For the affine image warping operator \( W(\cdot) \) defined in (1) and (2), the following is true:

\[
\mu(W(f)) = |\text{det}(M)| \mu(f). \tag{35}
\]

**Proof:** Consider the coordinates of original image \( f(x) \) and its warped version \( W(f)(u) \), which are related by (2). We have \( u = W(x) \), and therefore

\[
\mu(W(f)(u)) = \int_{u \in D_{W(i)}} du = \int_{x \in D_{\Delta^*}} |\text{det}(M)| dx
\]

This completes the proof of Theorem 2.
Theorem 2 shows that \( f(x) \) will be scaled up when \( \Delta' \) and scaled down when \( -\Delta' \).

**Corollary 2:** Assume \( f(x) \) can be precisely reconstructed from samples \( f[i] = D \Delta' f(x) \). If sampling \( f(x) \) with interval \( \Delta'/\sqrt{\det(M)} \) does not incur aliasing, then aliasing will not occur when sampling \( W(f(x)) \) with interval \( \Delta' \).

**Proof:** Presume that we obtain \( P \) points by sampling \( f(x) \) with interval \( \Delta' \). Then, by Theorem 2, we will get \( \det(M) P \) sampling points when sampling \( W(f(x)) \) with interval \( \Delta' \). If \( \det(M) < 1 \), this is equivalent to downsampling \( f(x) \) by a factor of \( 1/\sqrt{\det(M)} \) (in horizontal and vertical directions). Then, aliasing free is guaranteed by the Nyquist sampling theorem if \( \Delta'/\sqrt{\det(M)} \leq \Delta_{Nyquist} \) is satisfied.

**Corollary 2** tells that \( f(x) \) and \( W(f(x)) \) can be precisely reconstructed from their samples \( f[i] \) and \( g[h] \), respectively, if \( \det(M) \leq \Delta_{Nyquist} \). Furthermore, for the set of \( K \) LR images, uniformly sampling \( f(x) \) and \( W(f(x)) \) with interval \( \Delta' \) will be aliasing free if \( \max \{\Delta'/\sqrt{\det(M_k)}\} \leq \Delta_{Nyquist} \) and \( \min \{\sqrt{\det(M)}\} \geq 1 \), for \( 1 \leq k \leq K \).

Obviously, this condition can be satisfied by choosing the frame of the highest spatial resolution as the reference image among \( \{g_k[p]\}_{k=1}^K \). When one takes \( \Delta' = \min \{\sqrt{\det(M_k)}\}_{k=1}^K \), \( \Delta = \Delta' \). The variation of scales among consecutive frames in an image sequence is generally small, i.e., \( \sqrt{\det(M_k)} \approx 1 \).

**APPENDIX C**

**GAUSS–NEWTON SOLUTION TO** \( \min_{\{f, \theta\}} \Phi(f, \theta, h^{(1)}) \)

Starting from current estimates \( \hat{f}^{(t)}, \hat{\theta}^{(t)} \), the Gauss–Newton solution to \( \min_{\{f, \theta\}} \Phi(f, \theta, h^{(1)}) \) can be obtained by performing the updates: \( f^{(t+1)} = f^{(t)} + \delta f \) and \( \theta^{(t+1)} = \theta^{(t)} + \delta \theta \), where \( \delta \) is the solution to the following linear system:

\[
\mathcal{M} \delta = \beta \tag{36}
\]

where \( \mathcal{M} \) and \( \beta \) are shown in the equations at the top of the page.

The Jacobian \( J(f, \theta, h^{(1)}) \) is of the following block diagonal form:

\[
J(f, \theta, h^{(1)}) = \frac{\partial}{\partial \theta} \left[ B(\theta, h^{(1)}) f \right] = \text{diag} \left[ J_k(f, \theta_k, h^{(1)}) \right]_{k=1}^K \tag{37}
\]

where

\[
J_k(f, \theta_k, h^{(1)}) = \frac{\partial}{\partial \theta_k} [Dh_k^T(\theta_k) f] = Dh_k^T G_k P \tag{38}
\]

where \( G_k \) is assembled from the gradient values at positions \( \{x_i, y_i\}_{i=1}^N \) in gradient field \( \nabla f(x) \):

\[
G_k = [\partial^2 \nabla f(x_k)]_{i=1}^N \quad \text{and} \quad P = [P(x_{i,y})]_{i=1}^N \quad \text{is a constant matrix of size } 2m_{ij} \times 6 \text{ for each frame under affine motion.}
\]

\( J_k(f, \theta_k, h^{(1)}) \) can be efficiently computed by the basic image manipulations of intensity scaling, filtering, and decimation instead of explicit constructions of \( D, \mathcal{H}_k, G_k, \) and \( P \).

The computation of \( G_k \) is, in essence, the estimation of gradient field \( \nabla f(x) \), and accordingly, \( G_k \) can be obtained through more sophisticated and accurate interpolations of \( \nabla f(x) \), whereas the methods in [13], [39], and [42] only use the four-neighbor interpolation to derive the Jacobians.

**ACKNOWLEDGMENT**

The authors would like to thank Associate Editor Prof. K. Lam and three reviewers for their constructive comments and suggestions that helped improve the clarity of presentation of this paper.

**REFERENCES**


