

An Efficient Algorithm for Video Superresolution Based on a Sequential Model*

P. Héas[†], A. Drémeau[‡], and C. Herzet[†]

Abstract. In this work, we propose a novel procedure for video superresolution, that is, the recovery of a sequence of high-resolution images from its low-resolution counterpart. Our approach is based on a “sequential” model (i.e., each high-resolution frame is supposed to be a displaced version of the preceding one) and considers the use of sparsity-enforcing priors. Both the recovery of the high-resolution images and the motion fields relating them is tackled. This leads to a large-dimensional, nonconvex and nonsmooth problem. We propose an algorithmic framework to address the latter. Our approach relies on fast gradient evaluation methods and modern optimization techniques for nondifferentiable/nonconvex problems. Unlike some other previous works, we show that there exists a provably convergent method with a complexity linear in the problem dimensions. We assess the proposed optimization method on several video benchmarks and emphasize its good performance with respect to the state of the art.

Key words. sparse models, nonconvex optimization, optimal control, video superresolution

AMS subject classifications. 90C06, 90C26, 68T45, 65K10, 65F22, 62H35

DOI. 10.1137/15M1023956

1. Introduction. Superresolution (SR) aims at reconstructing high-resolution (HR) images from distorted low-resolution (LR) observations. This type of methodology dates back to the 1970s with the pioneering work of Gerchberg [26] and de Santis and Gori [16]. Since then, SR has been applied to a large variety of applicative domains, including infrared [28], medical [44], satellite, and aerial [40, 49] imaging. We refer the reader to [36] for a pretty comprehensive overview of the works dealing with SR.

One can distinguish between different setups in the domain of SR. “Single-frame” SR aims at computing an enhanced version of some HR image from the observation of *one* single LR image; see, e.g., [17, 29, 41]. On the other hand, the “multiframe” paradigm typically focuses on the recovery of *one* HR image by exploiting the observations of *several* LR frames; see, e.g., [21, 39, 23, 25, 35, 51, 32]. Finally, the “video” SR problem consists in estimating a *sequence* of HR images from the observations of their LR counterparts. We consider the latter paradigm in this paper.

From a conceptual point of view, a simple (but valid) solution to address video SR consists in applying single-frame or multiframe procedures on each frame of the HR sequence to recover. This strategy was, for example, considered in [21, 39, 23, 25, 35, 51, 32]. Nevertheless, this approach may fail in properly exploiting the strong temporal correlations existing between the

*Received by the editors June 1, 2015; accepted for publication (in revised form) January 29, 2016; published electronically May 3, 2016.

<http://www.siam.org/journals/siims/9-2/M102395.html>

[†]INRIA Centre Rennes - Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes, France (patrick.heas@inria.fr, cedric.herzet@inria.fr).

[‡]ENSTA Bretagne and Lab-STICC UMR 6825, 29806 Brest, France (angelique.dremeau@ensta-bretagne.fr).

(successive) frames of the HR sequence. Hence, procedures specifically dedicated to accounting for these dependencies have been proposed in the literature; see [48, 19, 20, 53, 60, 22, 27, 47, 46]. A central element is the “sequential” model linking the frames of the HR sequence. More specifically, in most of these methods, the frames of the HR sequence are supposed to obey a dynamical model where each HR image is seen as a displaced version (by some unknown motion field) of the preceding one (see section 3 for a detailed description). This is in contrast with the standard multiframe model where each LR observation is assumed to be an LR displaced version of one given reference frame.

The practical exploitation of the sequential model nevertheless faces a certain number of bottlenecks. The most stringent one is probably the model dimensionality: because it accounts for the temporal evolution of each HR frame, the number of variables involved in the sequential model may become very large. This makes video SR based on sequential models pretty challenging. As a matter of fact, in comparison with the huge number of papers dealing with SR, only a few have focused on this particular problem; see [48, 19, 20, 53, 60, 22, 27, 47, 46]. In [48], the authors modeled the dependence between the different images of the sequence as a Gaussian process and provided an efficient implementation in the Fourier domain. Other contributions relied on adaptive-filtering techniques; see [19, 20, 53, 60, 22, 27]. In this line of thought, most of the contributions cited above considered that the HR sequence is ruled by a state-space sequential model and the authors derived estimation procedures inspired by the well-known Kalman filter. The standard Kalman updates leading to a prohibitive complexity in the context of video SR, Elad and coauthors published a series of papers [19, 20, 22] in which they proposed updates having a linear complexity in the problem dimensions. Their approach is based on some approximations of the model and/or Kalman updates (e.g., uniform translational motion [22], noise-free evolution model [20], etc.). In [47, 46], the authors considered a local approximation of the state-space sequential model by using steering kernel regression on the LR observations.

In this paper, we provide an approximation-free methodological framework exploiting a sequential model for video SR. We express the unknown HR sequence as the solution of a constrained optimization problem and propose an iterative procedure to solve the latter. Our method is provably convergent (to a local minimum of the problem) and has a tractable complexity per iteration (i.e., linear in the problem dimensions). The proposed framework encompasses two important ingredients of video SR, namely, (i) a precise characterization of the motion fields linking the successive frames of the sequence and (ii) the exploitation of proper priors on the unknowns of the problem. These two ingredients lead to additional difficulties (on top of the large dimensionality) since they typically introduce *nonconvex* and *nonsmooth* terms in the cost function to minimize. We elaborate on these points in the next two paragraphs.

A precise characterization of the model connecting the different images of the HR sequence is crucial for the success of video SR. Typically, videos are characterized by nonglobal motions. This is in contrast with many standard SR models of the literature which assume global motions (e.g., translation [1], affine [56], or projection [12]), well-suited to still image reconstruction. The imaging model in video SR thus takes a more involved form and has to be considered with care. In particular, the estimation of the motion between two consecutive frames is usually tantamount to solving an optical-flow problem [4]. Embedding motion estimation in the SR reconstruction introduces new difficulties: (i) it increases the

problem dimensionality since two additional unknowns (the displacement in each direction) have to be estimated for each pixel of the HR images; (ii) it typically introduces nonlinearities in the image formation model. These obstacles are particularly prominent in the case of a sequential model because of the nested structure of the unknowns dependencies. As a consequence, until recently, motion estimation has been overlooked and considered as a side problem in many SR contributions involving either mutiframe or sequential models (see, e.g., [39, 21, 19, 20, 53, 60, 23, 22, 27, 51, 35]) with the exception of, e.g., [4, 25, 30, 32].

Interestingly, several authors have emphasized the importance of accurate motion estimation in the video SR process and provided studies of the sensibility of adaptive-filtering techniques to the latter; see [60, 13, 14, 15]. In this paper, we show that the motion estimation can be included in our video SR problem without significantly increasing the computational cost.

Another important ingredient for the success of video SR is the definition of proper priors on (some of) the unknowns of the problem. Indeed, video SR is a naturally ill-posed problem: typical setups impose the observation of (at most) one LR image per frame of the HR sequence; hence, if the motion between the different frames is unknown, it is easy to see that the number of variables which have to be estimated is well beyond the number of observations. In order to tackle this difficulty, a well-known technique consists in resorting to prior information on the sought quantities. This type of approach has been used extensively (but not only) in the context of single-image SR, where an HR image has to be reconstructed from one single LR observation. First methodologies based on prior information date back to the 1970s [26, 16]. Since then, many types of priors have been studied, including Markov random fields [45], total variation [35, 58, 32], morphological [42], or sparse [57, 41] models, etc. Among the most effective models in the literature, many rely on the minimization of some nondifferentiable functions. It is, for example, the case of SR techniques based on sparse representations where the decomposition coefficients of the sought quantity in a redundant dictionary are commonly penalized by an ℓ_1 norm; see, e.g., [24]. Another example is total variation where the ℓ_1 norm is applied to the gradient of the sought images/motions; see, e.g., [32]. The introduction of nondifferentiable functions in the SR reconstruction leads to new conundrums since standard optimization techniques for smooth problems can no longer be applied. As mentioned previously, we address this problem in the paper as well. Hereafter, we mainly focus on problems involving an ℓ_1 norm, although other nondifferentiable convex functions could be processed using a procedure similar to the one exposed in this paper.

In summary, in this paper we propose a methodological framework for video SR based on a sequential model. We consider the estimation of both a sequence of HR images and the motion fields relating them, while allowing for some nondifferential terms in the cost function. Our approach is based on the combination of several modern optimization tools: fast gradient computation [6], the “alternating direction method of multipliers” (ADMM) [7] for large-scale nondifferentiable convex problems, and a recent procedure for nonconvex and nondifferentiable optimization proposed by Attouch et al. in [2, 3]. The resulting algorithm is ensured to converge to a local minimum of the problem while having a linear complexity per iteration in the problem dimensions. We illustrate the good behavior of the proposed method with respect to other techniques of the state of the art in several setups.

The rest of the paper is organized as follows. We introduce the notations used throughout the paper in section 2. In section 3, we present the sequential model considered in our

subsequent derivations. In section 4, we express the video SR problem as a constrained optimization problem and provide a numerical procedure to solve it. The overall procedure is described in subsection 4.3 whereas two important algorithmic building blocks are presented in subsections 4.1 and 4.2. The numerical evaluation of the proposed method is carried out in section 5 for different experimental setups.

2. Notations. The notational conventions adopted in this paper are as follows. Italic lowercase indicates a scalar quantity, as in a ; boldface lowercase (resp., uppercase) indicates a vector (resp., matrix) quantity, as in \mathbf{a} (resp., \mathbf{A}). The n -dimensional vector of zeroes and identity matrix will be written as $\mathbf{0}_n$ and \mathbf{I}_n . The i th element of vector \mathbf{a} is denoted $\mathbf{a}(i)$; similarly $\mathbf{A}(i, j)$ is the element of \mathbf{A} located at row i and column j . The exponent $*$ denotes the transpose operation. A subscript notation, as in \mathbf{a}_t , will refer to the member of some sequence $\{\mathbf{a}_t\}_{t=0}^T = \{\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_T\}$.

Calligraphic letters, such as \mathcal{H} , denote functions. The subscript notation \mathcal{H}_i may either denote the i th element of a set $\{\mathcal{H}_i\}_i$ or the i th component of a multidimensional function $\mathcal{H} : \mathbb{R}^m \rightarrow \mathbb{R}^n$; the distinction between these two notations is usually clear from the context. The Jacobian matrix of $\mathcal{H} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ evaluated at $\tilde{\mathbf{a}}$, denoted by $\nabla_{\mathbf{a}}\mathcal{H}(\tilde{\mathbf{a}})$, is defined as

$$\nabla_{\mathbf{a}}\mathcal{H}(\tilde{\mathbf{a}}) = \begin{pmatrix} \partial_{\mathbf{a}(1)}\mathcal{H}_1(\tilde{\mathbf{a}}) & \cdots & \partial_{\mathbf{a}(m)}\mathcal{H}_1(\tilde{\mathbf{a}}) \\ \vdots & & \vdots \\ \partial_{\mathbf{a}(1)}\mathcal{H}_n(\tilde{\mathbf{a}}) & \cdots & \partial_{\mathbf{a}(m)}\mathcal{H}_n(\tilde{\mathbf{a}}) \end{pmatrix} \in \mathbb{R}^{n \times m},$$

where ∂_v is the partial derivative operator with respect to v . We use the notation $\nabla_{\mathbf{a}}\mathcal{H}^*(\tilde{\mathbf{a}})$ to denote the transpose of $\nabla_{\mathbf{a}}\mathcal{H}(\tilde{\mathbf{a}})$.

3. Model. Let $\mathbf{x}_t \in \mathbb{R}^n$ be the image at time t of an HR video sequence rearranged into an n -dimensional vector, with $t \in \{0, \dots, T\}$. Let us suppose that we capture noisy and LR observations $\mathbf{y}_t \in \mathbb{R}^m$ with $m \leq n$ of the HR sequence: $\forall t \in \{0, \dots, T\}$,

$$(3.1) \quad \mathbf{y}_t = \mathcal{H}(\mathbf{x}_t) + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t \in \mathbb{R}^m$ stands for some noise and $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ denotes a linear function, which is the composition of a low-pass filtering and a subsampling operation. We focus on the problem of recovering the HR sequence $\{\mathbf{x}_t\}_{t=0}^T$ from the LR observations $\{\mathbf{y}_t\}_{t=0}^T$.

Without any additional information, this problem is ill-conditioned since the number of unknowns (that is $(T+1)n$) is larger than the number of observations (that is $(T+1)m$). One way to circumvent this problem is to take into account the relation existing between the HR images at different time instants. More specifically, as part of a video, we can assume that two consecutive images obey the following sequential model:¹

$$(3.2) \quad \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1},$$

where $\mathcal{P} : \mathbb{R}^n \times \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ is a ‘‘warping’’ function characterized by a displacement $\mathbf{d}_{t+1} \in \mathbb{R}^{2n}$, and $\boldsymbol{\epsilon}_{t+1} \in \mathbb{R}^n$ is some noise. The choice of \mathcal{P} is usually motivated by some conservation

¹We note that *backward* sequential models such as (3.2) are common in the computer-vision literature. We therefore restrict our reasoning to the latter formulation. However, adapting the methodologies derived in section 4 to a *forward* sequential model is straightforward.

property, as for example the preservation of the pixel intensity along the displacement. One particular instance of a function \mathcal{P} , that we will consider in what follows, is based on the well-known “displaced frame difference” model. More specifically, this model assumes that the s th component of $\mathcal{P}(\mathbf{x}_t, \mathbf{d}_t)$ admits the following series representation:

$$(3.3) \quad \mathcal{P}_s(\mathbf{x}_t, \mathbf{d}_t) = \sum_{i \in \mathcal{V}(\chi(s) + \mathbf{d}_t(s))} \mathbf{x}_t(i) \psi_i(\chi(s) + \mathbf{d}_t(s)),$$

where $\chi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a function returning the spatial position corresponding to index s , $\mathcal{V}(\chi(s) + \mathbf{d}_t(s))$ denotes a subset of indices corresponding to the “neighborhood” of point $\chi(s) + \mathbf{d}_t(s)$, and $\{\psi_i\}_{i=1}^n$ with $\psi_i : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a family of bidimensional polynomial interpolation functions. In this case, (3.2)–(3.3) models the fact that \mathbf{x}_t can be seen as a displaced version of \mathbf{x}_{t+1} plus some additive noise. Let us note that \mathcal{P} , as defined in (3.3), is linear in \mathbf{x}_t and polynomial in \mathbf{d}_t ; it is thus a bipolynomial function. Let us also mention that \mathcal{V} typically only contains a few elements, that is $|\mathcal{V}| \ll n$, where $|\mathcal{V}|$ denotes the cardinality of \mathcal{V} ; this observation will play an important role in what follows for the analysis of the complexity of the proposed SR methodology.

The noise ϵ_{t+1} in (3.2) accounts for all the modifications of the image \mathbf{x}_t which cannot be inferred from \mathbf{x}_{t+1} and \mathbf{d}_{t+1} . This includes pixel occlusions, interpolation errors, or variations of the scene illumination. Notice that, in practice, the choice of \mathcal{P} should be made such that the residual noise ϵ_{t+1} is as small as possible. In particular, if $\epsilon_{t+1} = 0$ (and \mathbf{d}_{t+1} is known), \mathbf{x}_t is entirely determined from $\mathbf{x}_{t+1} \forall t$. Recovering the whole sequence $\{\mathbf{x}_t\}_{t=0}^T$ is then tantamount to recovering the last image \mathbf{x}_T . In such a case, the number of unknowns is therefore reduced to n and the recovery of the HR sequence from the LR images may be possible.

Another option to decrease the ill-posedness of the video SR problem consists in restricting the family of signals to which the “initial condition”² \mathbf{x}_T belongs. We will in particular³ consider the case where \mathbf{x}_T is assumed to be sparse in some (possibly redundant) dictionary $\mathbf{D} \in \mathbb{R}^{n \times q}$, that is

$$(3.4) \quad \mathbf{x}_T = \mathbf{D}\mathbf{c} \quad \text{for some } \mathbf{c} \in \mathbb{R}^q \text{ such that } \|\mathbf{c}\|_0 \ll n,$$

where $\|\cdot\|_0$ is the so-called “ ℓ_0 norm,” which returns the number of nonzero coefficients of its argument. Dealing with $\|\cdot\|_0$ leads to combinatorial optimization problems. Hereafter we will thus consider the ℓ_1 norm, which is a well-known surrogate to the ℓ_0 norm. In particular, if the sparsity of the sought vector is large enough, there exists an equivalence between the solution of the problems involving the ℓ_0 and ℓ_1 norms; see [24].

Finally, let us mention that the displacement \mathbf{d}_t between two successive images is rarely known in practice. It must therefore be inferred from the received LR images $\{\mathbf{y}_t\}_{t=0}^T$. This may seem to be counterproductive since the estimation of \mathbf{d}_t implies an increase of the number of unknowns of $2n$ elements per time step. One way to circumvent this problem consists

²We remind the reader that we consider a *backward* sequential model.

³The sparsity constraints could be imposed on every \mathbf{x}_t without introducing any conceptual problems in the methodology exposed in section 4.

again in constraining \mathbf{d}_t to belong to a restricted family of signals. In this paper, we consider an implicit restriction by enforcing some nonnegative function of \mathbf{d}_t to be small. More specifically, we assume that the sought displacement is such that $\mathcal{R}(\mathbf{G}^*\mathbf{d}_t)$ is “small,” where $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_h] \in \mathbb{R}^{2n \times h}$ is some linear “analysis” operator and $\mathcal{R} : \mathbb{R}^h \rightarrow \mathbb{R}$ is a nonnegative function. We note that this approach is commonly adopted in the computer-vision literature in which many options for \mathcal{R} and \mathbf{G} have been proposed; see [5]. This approach was also used in the “multiframe” setting [35, 32] where the motions between the reference HR frame and the LR observations were penalized to have a small total variation (TV) norm. In what follows, we will focus on the following choice for \mathbf{G} and \mathcal{R} : the elements of $\mathbf{G}^*\mathbf{d}_t$ will correspond to the spatial gradients of (an interpolation of) \mathbf{d}_t at each point of the pixel grid; $\mathcal{R}(\mathbf{G}^*\mathbf{d}_t)$ takes the form

$$(3.5) \quad \mathcal{R}(\mathbf{G}^*\mathbf{d}_t) \triangleq \sum_{i=1}^n \mathbf{w}(i) \left| \sum_{j \in \mathcal{S}_i} (\mathbf{g}_j^*\mathbf{d}_t)^2 \right|^{\frac{p}{2}}, \quad p \geq 0,$$

where \mathbf{w} is defined as a vector of weights and the \mathcal{S}_i 's denote disjoint subsets of elements of $\{1, \dots, h\}$. Index i represents a location on the pixel grid. The subset \mathcal{S}_i typically gathers 4 elements corresponding to the 2 spatial gradients of the 2 components of motion \mathbf{d}_t at the location indexed by i . For $p = 2$, these choices are equivalent to constraining the spatial gradient of the displacement field \mathbf{d}_t by a quadratic penalization [54], whereas the case $p = 1$ corresponds to the weighted TV approach suggested in [59].

In summary, (3.1)–(3.4) together with the definition of \mathbf{G} and \mathcal{R} specify our prior/observation model. In the next section, we will present a low-complexity methodology exploiting this model to recover the HR sequence from the collected observations $\{\mathbf{y}_t\}_{t=0}^T$. More specifically, we will assume that the unknowns of the problem include the HR sequence $\{\mathbf{x}_t\}_{t=0}^T$, the sequential noise $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$, the displacements $\{\mathbf{d}_t\}_{t=1}^T$, and the decomposition vector \mathbf{c} . All the other parameters of the problem will be supposed to be known, although they could easily be included as additional unknowns without introducing any conceptual problem in the proposed methodology.

4. The estimation procedure. In this section, we expose our methodology to estimate the HR sequence by exploiting the model described in section 3. Our approach is based on the resolution of a constrained optimization problem. We introduce the following shorthand notations: $\mathbf{x} \triangleq \{\mathbf{x}_t\}_{t=0}^T$, $\boldsymbol{\epsilon} \triangleq \{\boldsymbol{\epsilon}_t\}_{t=1}^T$, and $\mathbf{d} \triangleq \{\mathbf{d}_t\}_{t=1}^T$. Our SR reconstruction procedure relies on the following constrained optimization problem:

$$(4.1) \quad \arg \min_{(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})} \mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) \quad \text{s.t.} \quad \begin{cases} \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}, & 0 \leq t \leq T-1, \\ \mathbf{x}_T = \mathbf{D}\mathbf{c}, \end{cases}$$

where

$$\mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) \triangleq \sum_{t=0}^T \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \alpha_1 \sum_{t=1}^T \|\boldsymbol{\epsilon}_t\|_p^p + \alpha_2 \sum_{t=1}^T \mathcal{R}(\mathbf{G}^*\mathbf{d}_t) + \alpha_3 \|\mathbf{c}\|_p^p$$

for some $\alpha_j \geq 0$, $j \in \{1, 2, 3\}$, and $p \geq 0$. Let us make a few comments about (4.1). The first constraint ensures that the images of the HR sequence verify the sequential model (3.2); the second enforces that prior model (3.4) is satisfied. Each term in the cost function $\mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$

has a clear physical meaning: the first term penalizes the discrepancies between the predicted and the received observations; the second penalizes the noise on the sequential model; the third enforces that the displacement has some regularity; and the last one constrains \mathbf{c} to have some desirable properties (depending on the choice of p). For example, setting $p \in [0, 1]$ typically promotes the sparsity of \mathbf{c} ; see [24]. Because sparsity has been revealed to be a good prior in a number of works, in the following our main objective is to find a solution to (4.1) with $p = 1$.

Problem (4.1) involves a huge number of unknowns (namely, $(4T + 1)n + q$ variables if \mathbf{x} , $\boldsymbol{\epsilon}$, \mathbf{d} , \mathbf{c} have to be estimated). Hence, solving (4.1) may be critical even for reasonable problem sizes: for instance, considering images of $n = 2^8 \times 2^8$ pixels, a nonredundant dictionary, i.e., $q = n$, and a sequence length $T = 2^4$, we have that the number of variables involved in the optimization problem grows up to roughly 2^{22} . Clearly, such a high-dimensional problem can only be addressed by specifically dedicated procedures. In subsection 4.3, we propose an overall methodology to solve (4.1) efficiently with $p = 1$. Our approach is based on the combination of several modern optimization tools, described in subsections 4.1 and 4.2. More specifically, the building blocks presented in subsections 4.1 and 4.2 tackle simplified versions of problem (4.1), which appear as intermediate steps in the overall procedure described in section 4.3. We briefly comment on these intermediate problems in the next paragraphs.

In section 4.1, we consider the case where $p = 2$ in (4.1), that is all the functions are differentiable. In such a case, we show that the gradient of the cost function associated with an (equivalent) unconstrained version of (4.1) can be evaluated efficiently by resorting to optimal control techniques [6]. More specifically, we emphasize that the complexity associated with the evaluation of the gradient of the cost function remains *linear* in the problem dimensions, for many setups of practical interest.

In section 4.2, we focus on the case where \mathbf{d} is known but $p = 1$. The corresponding optimization problem is then convex but not differentiable. Building on our derivations in section 4.1, we emphasize that this type of problem can be nicely addressed by resorting to the so-called ADMM [7], a modern optimization technique proposed to handle large-scale nondifferentiable optimization problems.

Finally, in section 4.3, we consider the general problem (4.1), where \mathbf{x} , $\boldsymbol{\epsilon}$, \mathbf{d} , \mathbf{c} have to be estimated and $p = 1$. In this case, (4.1) is nonconvex (because the term $\mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1})$ appearing in the constraints is bipolynomial) and nondifferentiable. In order to address this problem, we resort to an optimization procedure introduced by Attouch et al. [2] and Attouch, Bolte, and Svaiter [3] and particularized in [43] to multiframe SR. The procedure is iterative and exploits the building blocks derived in sections 4.1 and 4.2 to solve intermediate problems. The complexity per iteration is linear in the problem dimensions. Moreover, from the arguments exposed in [43], it can be shown that the proposed procedure is convergent to a critical point of the problem.

4.1. The first building block. In this section, we assume that $p = 2$ (so that all the functions appearing in (4.1) are differentiable) and show that an efficient resolution of (4.1) via gradient descent algorithms exists. Our approach is based on fast gradient evaluation techniques as exposed in [6].

In order to present our methodology, we first reformulate (4.1) as an (equivalent) unconstrained problem. Notice that, because of the constraints in problem (4.1), any \mathbf{x}_t can be

expressed as a deterministic function of \mathbf{c} and $\{\boldsymbol{\epsilon}_{t'}, \mathbf{d}_{t'}\}_{t'=t+1}^T$. In other words, there exists a function $\mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) : \mathbb{R}^{Tn} \times \mathbb{R}^{2Tn} \times \mathbb{R}^q \rightarrow \mathbb{R}^{(T+1)n}$ such that, given $\boldsymbol{\epsilon}, \mathbf{d}$, and \mathbf{c} , $\mathbf{x} = \mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ is the unique vector satisfying the constraints in (4.1). As a consequence, (4.1) can also be equivalently expressed as

$$(4.2) \quad \arg \min_{(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})} \mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}),$$

where

$$(4.3) \quad \begin{aligned} \mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) &\triangleq \mathcal{J}(\mathbf{x} = \mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}), \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}), \\ &= \sum_{t=0}^T \|\mathcal{H}(\mathcal{Q}_t(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})) - \mathbf{y}_t\|_2^2 + \alpha_1 \sum_{t=1}^T \|\boldsymbol{\epsilon}_t\|_p^p \\ &\quad + \alpha_2 \sum_{t=1}^T \mathcal{R}(\mathbf{G}^* \mathbf{d}_t) + \alpha_3 \|\mathbf{c}\|_p^p, \end{aligned}$$

and $\mathcal{Q}_t(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ is the restriction of $\mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ to \mathbf{x}_t .

Since $p = 2$, (4.2) is a smooth *unconstrained* minimization problem and can thus be solved by any procedure belonging to the family of gradient descent algorithms. At this point, let us make two remarks: (i) $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ usually has an intricate structure and its gradient does therefore not have any simple analytical expression; (ii) the computation of the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ via finite differences is out of reach for the considered problem because it would require us to evaluate the cost function twice as many times as the (huge!) number of variables.

As a consequence, the main bottleneck for solving (4.2) lies in the tractable evaluation of the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$. We emphasize in Appendix A that the particular structure of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ enables the use of a specific methodology with a complexity scaling linearly with the problem dimensions. More specifically, let

$$(4.4) \quad \begin{cases} \mathcal{G}_0(\mathbf{x}_0) \triangleq \|\mathcal{H}(\mathbf{x}_0) - \mathbf{y}_0\|_2^2, \\ \mathcal{G}_t(\mathbf{x}_t, \boldsymbol{\epsilon}_t, \mathbf{d}_t) \triangleq \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \alpha_1 \|\boldsymbol{\epsilon}_t\|_p^p + \alpha_2 \mathcal{R}(\mathbf{G}^* \mathbf{d}_t) \quad \text{for } 1 \leq t \leq T-1, \\ \mathcal{G}_T(\mathbf{x}_T, \boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}) \triangleq \|\mathcal{H}(\mathbf{x}_T) - \mathbf{y}_T\|_2^2 + \alpha_1 \|\boldsymbol{\epsilon}_T\|_p^p + \alpha_2 \mathcal{R}(\mathbf{G}^* \mathbf{d}_T) + \alpha_3 \|\mathbf{c}\|_p^p. \end{cases}$$

Using the notation $\mathcal{G}_T(\boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}) \triangleq \mathcal{G}_T(\mathbf{x}_T = \mathbf{D}\mathbf{c}, \boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c})$, the elements of the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ at $(\boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}')$ can then be evaluated as follows:

$$(4.5) \quad \begin{cases} \nabla_{\mathbf{d}_t} \mathcal{J}(\boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \nabla_{\mathbf{d}_t} \mathcal{P}^*(\mathbf{x}'_t, \mathbf{d}'_t) \boldsymbol{\zeta}_{t-1} + \nabla_{\mathbf{d}_t} \mathcal{G}_t(\mathbf{x}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t), \\ \nabla_{\boldsymbol{\epsilon}_t} \mathcal{J}(\boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \boldsymbol{\zeta}_{t-1} + \nabla_{\boldsymbol{\epsilon}_t} \mathcal{G}_t(\mathbf{x}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t), \\ \nabla_{\mathbf{c}} \mathcal{J}(\boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \mathbf{D}^* \boldsymbol{\zeta}_T + \nabla_{\mathbf{c}} \mathcal{G}_T(\boldsymbol{\epsilon}'_T, \mathbf{d}'_T, \mathbf{c}'), \end{cases}$$

where the variables $\mathbf{x}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t$, and \mathbf{c}' must satisfy the constraints of problem (4.1), that is

$$(4.6) \quad \begin{cases} \mathbf{x}'_T = \mathbf{D}\mathbf{c}', \\ \mathbf{x}'_t = \mathcal{P}(\mathbf{x}'_{t+1}, \mathbf{d}'_{t+1}) + \boldsymbol{\epsilon}'_{t+1}, \quad t = T-1, \dots, 0, \end{cases}$$

and the sequence of “adjoint” variables $\{\zeta_t\}_{t=0}^T$ obeys the following recursion:

$$(4.7) \quad \begin{cases} \zeta_0 = \nabla \mathcal{G}_0(\mathbf{x}'_0), \\ \zeta_t = \nabla_{\mathbf{x}_t} \mathcal{P}^*(\mathbf{x}'_t, \mathbf{d}'_t) \zeta_{t-1} + \nabla_{\mathbf{x}_t} \mathcal{G}_t(\mathbf{x}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t), \quad t = 1, \dots, T-1, \\ \zeta_T = \nabla_{\mathbf{x}_T} \mathcal{P}^*(\mathbf{x}'_T, \mathbf{d}'_T) \zeta_{T-1} + \nabla_{\mathbf{x}_T} \mathcal{G}_T(\mathbf{x}'_T, \boldsymbol{\epsilon}'_T, \mathbf{d}'_T, \mathbf{c}'). \end{cases}$$

Expressions in (4.5) together with recursions (4.6) and (4.7) provide an efficient way to evaluate the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$. The overall methodology can be understood as a 3-step procedure: (i) given some values of $\boldsymbol{\epsilon}'$, \mathbf{d}' , and \mathbf{c}' , evaluate $\{\mathbf{x}'_t\}_{t=0}^T$ with recursion (4.6); (ii) use the value of $\{\mathbf{x}'_t\}_{t=0}^T$ to evaluate the adjoint variables $\{\zeta_t\}_{t=0}^T$ from (4.7); (iii) compute the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ by using (4.5). Note that the gradients appearing in the right-hand side of (4.5) and (4.7) typically have simple analytical expressions and are thus straightforward to evaluate.

It is easy to see that the complexity induced by this methodology scales (at worst) as $\mathcal{O}(n^2T + nq)$ since it only involves matrix-vector multiplications, with matrices of dimension $n \times n$ or $n \times q$. In practice, this complexity can usually be reduced to $\mathcal{O}(nT + q)$, or simply to $\mathcal{O}(nT)$ in the case of a nonredundant dictionary. This linearity in the problem dimensions occurs if the matrices involved in (3.4), (4.5), and (4.7) are typically very sparse and/or rely on fast transforms of linear complexity.⁴ In the (typical) example (3.3) considered in this paper, we clearly obtain this linear complexity since $|\mathcal{V}(\chi(s) + \mathbf{d}_t(s))| \ll n$. In the rest of the paper, we focus on model (3.3) and choose a dictionary so that the complexity related to (4.5)–(4.7) is linear in the problem dimensions.

Before concluding this section, let us make a remark to highlight some connections with some previous works which considered the “Kalman smoother” update rules as the starting point of their video SR method; see [19]. First notice that, assuming \mathbf{d} is known, (4.2) with $p = 2$ corresponds to the “maximum a posteriori” (MAP) estimation problem associated with the following probabilistic (backward) state-evolution model:

$$(4.8) \quad \begin{cases} \mathbf{x}_T \sim \mathcal{N}(\mathbf{0}_n, \alpha_3^{-1} \mathbf{D} \mathbf{D}^*), \\ \mathbf{x}_t \sim \mathcal{N}(\mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}), \alpha_1^{-1} \mathbf{I}_n), \\ \mathbf{y}_t \sim \mathcal{N}(\mathcal{H}(\mathbf{x}_t), \mathbf{I}_m), \end{cases}$$

where $\mathbf{v} \sim \mathcal{N}(\mathbf{m}, \Gamma)$ indicates that \mathbf{v} is distributed according to a multivariate normal distribution with mean \mathbf{m} and covariance Γ .

For such a model, it is well known that the Kalman smoother can compute exactly the solution of (4.2) in a finite number of steps, namely, one forward and one backward recursions; see, e.g., [34, Chapter 20]. The Kalman smoother involves the update of a length- n mean vector and an $n \times n$ covariance matrix at each step of the two recursions; moreover, the evaluation of these quantities requires the inversion of an $n \times n$ matrix. Hence, the Kalman

⁴This is the case for any nonredundant wavelet basis, which will induce an overall complexity of $\mathcal{O}(nT)$. Fast transforms for sparse redundant dictionaries such as curvelets frames also exist but imply a slight complexity overload since the matrix-vector multiplication scales in this case as $\mathcal{O}(n \log n)$, yielding an overall complexity of $\mathcal{O}(n(T + \log n))$.

smoother exhibits a computational complexity scaling as⁵ $\mathcal{O}(n^3T)$. Since this complexity is prohibitive for most practical setups, several approximations of the Kalman updates for video SR have been proposed in [19]. On the other hand, the procedure described in this section provides an alternative, approximation-free, solution to the MAP problem. Indeed, since (4.2) with $p = 2$ is a differentiable problem, it can be solved with a simple gradient descent method. More specifically, we can apply the methodology described in this section to efficiently compute the gradient of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ with respect to $\boldsymbol{\epsilon}$ and \mathbf{c} (using the two last rows of (4.5)). The complexity of this method then only scales as $\mathcal{O}(nT)$ per iteration. Moreover, because $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ is strictly convex in $(\boldsymbol{\epsilon}, \mathbf{c})$, this type of algorithm is ensured to converge to the global minimum of the problem. Hence, if the descent algorithm has converged (close) to the minimum after a reasonable number of iterations, the proposed methodology drastically reduces the complexity necessary to obtain the MAP solution as compared to a Kalman smoother.

4.2. The second building block. In this section, we address problem (4.1) with $p = 1$ but assume that \mathbf{d} is known (and thus therefore no longer appears as an optimization variable in (4.1)). Particularizing (4.1) to these working hypotheses, we obtain the following convex but nondifferentiable problem:

$$(4.9) \quad \arg \min_{(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c})} \mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \quad \text{s.t.} \quad \begin{cases} \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}, & 0 \leq t \leq T-1, \\ \mathbf{x}_T = \mathbf{D}\mathbf{c}, \end{cases}$$

where

$$\mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \triangleq \sum_{t=0}^T \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \alpha_1 \sum_{t=1}^T \|\boldsymbol{\epsilon}_t\|_1 + \alpha_3 \|\mathbf{c}\|_1.$$

This problem is convex but nondifferentiable. As previously, the main bottleneck for its resolution lies in its high dimensionality. This, in turn, forces us to resort to low-complexity optimization procedures. We show hereafter, that a complexity scaling linearly with the problem dimensions is possible by using the ADMM. ADMM has recently emerged in the optimization community as a method to address large-scale optimization problems. Among the particular assets of this type of method, let us mention (i) its robustness (the convergence to a global minimum is ensured under very mild conditions); (ii) its rapid convergence to an acceptable accuracy (typically a few tens of iterations is sufficient). We refer the reader to Appendix B for a short description of the ADMM framework.

In order to derive the ADMM recursions, we first need to reformulate (4.9) in the standard form (B.1) in Appendix B. Letting

$$(4.10) \quad \Omega \triangleq \left\{ (\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \left| \begin{array}{l} \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}, \quad 0 \leq t \leq T-1 \\ \mathbf{x}_T = \mathbf{D}\mathbf{c} \end{array} \right. \right\},$$

⁵We note that the complexity can be reduced to $\mathcal{O}(T(m^3 + mn))$ by using some computational tricks such as the well-known Woodbury matrix identity; see, e.g., [34, Lemma 4.1]. However, the latter still remain too costly for typical problem sizes.

(4.9) can be reexpressed as

$$\begin{aligned} \arg \min_{\substack{(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \in \Omega \\ (\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{c}}) \in \mathbb{R}^{nT+q}}} & \sum_{t=0}^T \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \alpha_1 \sum_{t=1}^T \|\tilde{\boldsymbol{\epsilon}}_t\|_1 + \alpha_3 \|\tilde{\mathbf{c}}\|_1 \\ \text{s.t.} & \begin{cases} \boldsymbol{\epsilon} = \tilde{\boldsymbol{\epsilon}}, \\ \mathbf{c} = \tilde{\mathbf{c}}. \end{cases} \end{aligned}$$

Here, we have added two new variables to the problem, $\tilde{\boldsymbol{\epsilon}}$ and $\tilde{\mathbf{c}}$, which are counterbalanced by the inclusion of two new constraints. Using the formalism exposed in Appendix B with $\mathbf{z}_1 = (\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c})$, $\mathbf{z}_2 = (\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{c}})$, $\Xi_1 = \Omega$, and $\Xi_2 = \mathbb{R}^{nT+q}$, we obtain the following ADMM recursions:

$$(4.11) \quad (\mathbf{x}^{(k+1)}, \boldsymbol{\epsilon}^{(k+1)}, \mathbf{c}^{(k+1)}) = \arg \min_{\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}} \mathcal{L}^{(k)}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c})$$

$$(4.12) \quad \begin{aligned} \text{s.t.} & \begin{cases} \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}, & 0 \leq t \leq T-1, \\ \mathbf{x}_T = \mathbf{D}\mathbf{c}, \end{cases} \\ & \begin{cases} \tilde{\boldsymbol{\epsilon}}_t^{(k+1)} = \arg \min_{\tilde{\boldsymbol{\epsilon}}_t} \|\tilde{\boldsymbol{\epsilon}}_t\|_1 + \frac{\rho_1}{2\alpha_1} \|\boldsymbol{\epsilon}_t^{(k+1)} - \tilde{\boldsymbol{\epsilon}}_t + \mathbf{u}_{\tilde{\boldsymbol{\epsilon}}_t}^{(k)}\|_2^2, \\ \tilde{\mathbf{c}}^{(k+1)} = \arg \min_{\tilde{\mathbf{c}}} \|\tilde{\mathbf{c}}\|_1 + \frac{\rho_3}{2\alpha_3} \|\mathbf{c}^{(k+1)} - \tilde{\mathbf{c}} + \mathbf{u}_{\tilde{\mathbf{c}}}^{(k)}\|_2^2, \end{cases} \end{aligned}$$

$$(4.13) \quad \begin{cases} \mathbf{u}_{\tilde{\boldsymbol{\epsilon}}_t}^{(k+1)} = \mathbf{u}_{\tilde{\boldsymbol{\epsilon}}_t}^{(k)} + \boldsymbol{\epsilon}_t^{(k+1)} - \tilde{\boldsymbol{\epsilon}}_t^{(k+1)}, \\ \mathbf{u}_{\tilde{\mathbf{c}}}^{(k+1)} = \mathbf{u}_{\tilde{\mathbf{c}}}^{(k)} + \mathbf{c}^{(k+1)} - \tilde{\mathbf{c}}^{(k+1)}, \end{cases}$$

where $\rho_1, \rho_3 > 0$ and we have introduced the function

$$(4.14) \quad \mathcal{L}^{(k)}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) = \sum_{t=0}^T \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \frac{\rho_1}{2} \|\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{(k)} + \mathbf{u}_{\tilde{\boldsymbol{\epsilon}}}^{(k)}\|_2^2 + \frac{\rho_3}{2} \|\mathbf{c} - \tilde{\mathbf{c}}^{(k)} + \mathbf{u}_{\tilde{\mathbf{c}}}^{(k)}\|_2^2.$$

Equations (4.11), (4.12), and (4.13) correspond, respectively, to expressions (B.2), (B.3), and (B.4) in Appendix B.

Let us make the following remarks about the different steps of the ADMM procedure. First, problem (4.11) has the same structural form as the problem addressed in section 4.1; in particular, all the terms of the cost function appearing in (4.11) are differentiable while the set of constraints imposed on $\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}$, and \mathbf{c} is strictly the same. We can thus apply the methodology described in section 4.1 to solve this problem via a gradient descent algorithm, with a complexity per iteration scaling as $\mathcal{O}(nT)$. Interestingly, let us mention that, under very mild conditions, the convergence of ADMM is still guaranteed if the minimizations in (4.11)–(4.12) are not performed exactly; see, e.g., [18, Theorem 8]. This suggests that the number of gradient steps carried out to search for the minimum of (4.11) can be rather limited without affecting the convergence of the overall ADMM process.

Second, the optimization problems specified in (4.12) have a very simple analytical solution. In fact the right-hand sides of (4.12) correspond to the definition of the proximal operator of the ℓ_1 norm. The latter has been extensively studied in the literature (see, e.g., [38, section 6.5.2]) and possesses a simple analytical solution based on soft-thresholding operators.

In particular, we have

$$(4.15) \quad \begin{cases} \tilde{\boldsymbol{\epsilon}}_t^{(k+1)}(i) = \text{soft}_{\frac{\alpha_1}{\rho_1}} \left(\boldsymbol{\epsilon}_t^{(k+1)}(i) + \mathbf{u}_{\boldsymbol{\epsilon}_t}^{(k)}(i) \right) \\ \tilde{\mathbf{c}}^{(k+1)}(i) = \text{soft}_{\frac{\alpha_3}{\rho_3}} \left(\mathbf{c}^{(k+1)}(i) + \mathbf{u}_{\mathbf{c}}^{(k)}(i) \right) \end{cases} \quad \forall i,$$

where

$$(4.16) \quad \text{soft}_\lambda(a) = \begin{cases} a - \lambda & \text{if } a \geq \lambda, \\ a + \lambda & \text{if } a \leq -\lambda, \\ 0 & \text{otherwise.} \end{cases}$$

We note that the solution of (4.12) is typically sparse since the soft-thresholding operator (4.16) enforces that the small coefficients are equal to zero. Moreover, we see from (4.15) that the complexity of this ADMM step clearly scales as $\mathcal{O}(nT + q)$.

As a conclusion, since the last step (4.13) of the procedure only involves vector additions, the particularization of ADMM to our problem leads to an algorithm exhibiting a complexity per iteration scaling linearly in the problem dimensions.

4.3. The overall procedure. Let us now concentrate our attention on our target problem, that is (4.1) with $p = 1$, where all the variables \mathbf{x} , \mathbf{c} , $\boldsymbol{\epsilon}$, \mathbf{d} have to be estimated. The cost function then contains both nondifferentiable and nonconvex terms. In such a case, ensuring the convergence to a global minimum is usually out of reach for any deterministic optimization procedure. In this section, we consider an optimization method proposed in [2, 3] and particularized to multiframe SR problems in [43]. This procedure addresses optimization problems involving a cost function satisfying the so-called ‘‘Kurdyca–Lojasiewicz’’ property and is guaranteed to converge to a critical point of the latter under mild conditions. We refer the reader to [2, 3] for more details about Kurdyca–Lojasiewicz functions. Here, we just mention that functions made up of the composition of piecewise polynomial functions obey the Kurdyca–Lojasiewicz property. Scrutinizing the structure of (4.1) and taking (3.3) into account, it is easy to see that our cost function is piecewise polynomial; the optimization framework developed in [2, 3] therefore applies.

Our methodology obeys a 2-step recursion which follows the same lines as the procedure presented in [43]. The building blocks described in subsections 4.1 and 4.2 are used to provide an efficient implementation of the intermediate problems appearing in these two steps. To express the procedure recursions, we focus on the unconstrained formulation (4.2) (with $p = 1$) of our general optimization problem (4.1). The first step of the procedure solves the following problem:

$$(4.17) \quad (\boldsymbol{\epsilon}^{(k+1)}, \mathbf{c}^{(k+1)}) = \arg \min_{(\boldsymbol{\epsilon}, \mathbf{c})} \mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}^{(k)}, \mathbf{c}) + \gamma \mathcal{C}(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^{(k)}, \mathbf{c} - \mathbf{c}^{(k)}),$$

where $\gamma > 0$, \mathcal{J} is the cost function in (4.3) with $p = 1$, and $\mathcal{C} : \mathbb{R}^{nT} \times \mathbb{R}^q \rightarrow \mathbb{R}_+$ is a nonnegative proper lower-semicontinuous convex function such that $\mathcal{C}(\mathbf{0}_{nT}, \mathbf{0}_q) = 0$. It thus consists in minimizing the (penalized) cost function $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ over the subset of variables $(\boldsymbol{\epsilon}, \mathbf{c})$; the penalizing term \mathcal{C} plays the role of a ‘‘cost-to-move’’ function which prevents the new iterate $(\boldsymbol{\epsilon}^{(k+1)}, \mathbf{c}^{(k+1)})$ from differing too much from the previous one. In what follows,

we will focus on the following penalizing term:⁶

$$(4.18) \quad \mathcal{C}(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^{(k)}, \mathbf{c} - \mathbf{c}^{(k)}) = \sum_{t=1}^T \|\mathbf{H}^*(\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t^{(k)})\|_1 + \|\mathbf{c} - \mathbf{c}^{(k)}\|_1,$$

where $\mathbf{H} \in \mathbb{R}^{n \times n}$ is a wavelet basis. The operational meaning of this cost-to-move function is as follows: the ℓ_1 norm enforces its argument to be sparse; hence, the second term in (4.18) ensures that the number of nonzero coefficients in $\mathbf{c}^{(k+1)}$ does not differ too much from the one in $\mathbf{c}^{(k)}$, while the first term plays the same role for the wavelet coefficients of $\boldsymbol{\epsilon}_t^{(k+1)} - \boldsymbol{\epsilon}_t^{(k)}$. Using this type of cost-to-move is not mandatory for the convergence of the proposed procedure. However, it has been shown empirically in [43] that it is well-suited to avoid some undesirable local minima of the cost function.⁷

In the second step of the recursion, we update the velocity field \mathbf{d} as

$$(4.19) \quad \mathbf{d}^{(k+1)} = \arg \min_{\mathbf{d}} \mathcal{B}(\mathbf{d}, \mathbf{d}^{(k)}) + \alpha_2 \sum_{t=1}^T \mathcal{R}(\mathbf{G}^* \mathbf{d}_t),$$

where $\mathcal{B}(\mathbf{d}, \mathbf{d}^{(k)})$ is a quadratic approximation⁸ of

$$\mathcal{B}(\mathbf{d}) \triangleq \sum_{t=0}^T \|\mathcal{H}(\mathcal{Q}_t(\boldsymbol{\epsilon}^{(k+1)}, \mathbf{d}, \mathbf{c}^{(k+1)})) - \mathbf{y}_t\|_2^2,$$

that is

$$(4.20) \quad \mathcal{B}(\mathbf{d}, \mathbf{d}^{(k)}) \triangleq \mathcal{B}(\mathbf{d}^{(k)}) + \nabla_{\mathbf{d}} \mathcal{B}^*(\mathbf{d}^{(k)})(\mathbf{d} - \mathbf{d}^{(k)}) + \frac{\alpha^{(k)}}{2} \|\mathbf{d} - \mathbf{d}^{(k)}\|_2^2, \quad \alpha^{(k)} > 0.$$

The choice of $\alpha^{(k)}$ is of course not arbitrary and should be made so that the convergence of the procedure is ensured. We elaborate on this point further in this section. For now, let us first discuss the practical implementation and complexity of recursions (4.17)–(4.19). It should be noticed that the building blocks presented in sections 4.1 and 4.2 can be exploited to solve these steps efficiently. Indeed, problem (4.17) has the same structural form as the one considered in (4.9): the cost function consists in a quadratic term plus a set of convex but nondifferentiable terms. We can thus use the ADMM procedure described in section 4.2 to address it. In the same way, we see from definition (4.20) that the cost function (4.19) is made up of a quadratic term plus some nondifferentiable function $\alpha_2 \sum_{t=1}^T \mathcal{R}(\mathbf{G}^* \mathbf{d}_t)$. Hence, the ADMM procedure described in section 4.2 can also be applied here to solve (4.19). In comparison to our exposition in section 4.2, only the proximal operators of the nondifferentiable terms will change when ADMM is applied to (4.17) and (4.19). In particular, the computation of any gradient of the differentiable part of the cost functions can be efficiently evaluated via the procedure described in section 4.1. We particularize the expression of the proximal operators appearing in the ADMM implementation of (4.17)–(4.19) in Appendix C.2. As previously,

⁶In theory, the ℓ_1 norm should be substituted for by a smooth approximation to prove convergence towards a critical point of the cost function, as was done in [43]. In practice, we note that this substitution does not impact convergence.

⁷An intuitive explanation is that the cost-to-move (4.18) induces a “coarse-to-grain” refinement of the unknowns which is usually beneficial in computer-vision problems; see details in [43].

⁸Note that $\mathcal{B}(\mathbf{d})$ is similar to the first term of the cost function in (4.3).

it turns out that the implementation of the latter only requires a linear complexity. The complexity of each iteration of (4.17), (4.19) is thus once again linear.

To conclude this section, let us discuss the convergence of the proposed procedure. In [43, Theorem 1], the authors proved that if $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ satisfies the Kurdyca–Lojasiewicz property and the $\alpha^{(k)}$'s are properly selected, the sequence defined in (4.17)–(4.19) is either unbounded or converges to a critical point of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$. A procedure to properly select factors $\alpha^{(k)}$ is exposed in [43, section 2.3] and is easy to implement in practice. Particularized to the setup considered in this paper, this procedure reads as follows: select $\alpha^{(k)} = 2^i \xi$ with $\xi > 0$ and with i the smallest positive integer such that

$$(4.21) \quad \begin{aligned} \mathcal{B}(\mathbf{d}^{(k+1)}) - \mathcal{B}(\mathbf{d}^{(k)}) &\leq \frac{(2^i - 1)\xi}{2} \|\mathbf{d}^{(k+1)} - \mathbf{d}^{(k)}\|^2 + \nabla_{\mathbf{d}} \mathcal{B}^*(\mathbf{d}^{(k)})(\mathbf{d}^{(k+1)} - \mathbf{d}^{(k)}) \\ &+ \alpha_2 \sum_{t=1}^T \left(\mathcal{R}(\mathbf{G}^* \mathbf{d}_t^{(k)}) - \mathcal{R}(\mathbf{G}^* \mathbf{d}_t^{(k+1)}) \right). \end{aligned}$$

As mentioned at the beginning of the section, the cost function $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ is piecewise polynomial and therefore satisfies the Kurdyca–Lojasiewicz property. Hence, the sequence defined by (4.17), (4.19) with factor selection (4.21) is either unbounded or converges to a critical point of $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$. Finally, let us note that the boundedness of $\{(\boldsymbol{\epsilon}^{(k)}, \mathbf{d}^{(k)}, \mathbf{c}^{(k)})\}_k$ is usually observed in practice or is easy to enforce by adding box constraints to the optimization problem.

5. Experiments. In this section, we provide an experimental validation of the SR procedure proposed in section 4.3. We focus on the problem of recovering a sequence of HR natural images from blurry and LR observations. In section 5.1, we provide a precise definition of the model parameters used to run our algorithms. In section 5.2, we describe several algorithms of the state of the art which will serve as points of comparison with the proposed approach. In sections 5.3 and 5.4, we respectively describe the databases and the figures of merit which will be used in our experiments. Finally, a discussion of the performance of the proposed SR methodology is provided in section 5.5.

5.1. Specification of the model and algorithm parameters. We first discuss the choice of the parameters appearing in the model described in section 3. In particular, we specify the definitions of \mathcal{H} , \mathcal{P} , \mathbf{D} , \mathbf{G} , and \mathcal{R} . We then provide some details about the parameters used in our algorithm.

The observation model \mathcal{H} is defined as the composition of a low-pass filtering and down-sampling operation. The low-pass filter is assumed to model the blurring effect induced by the camera transfer function. In our simulations, we use an approximation of a Gaussian kernel with a standard deviation equal to 1.12, as proposed in [8]. A down-sampling factor equal to 2 is considered.

The operator \mathcal{P} is supposed to model a “displaced frame difference”: \mathcal{P} is thus defined as in (3.3) with the interpolation functions $\{\psi_i\}_{i=1}^n$ equal to bidimensional cubic cardinal splines [52]. This representation offers a reasonable accuracy with a complexity scaling linearly with the image dimension; see Appendix C.1 for further details.

The dictionary \mathbf{D} is chosen so that natural images have a sparse representation as a combination of a few of its columns. Several choices of such dictionaries have been proposed in the literature; see, e.g., [33, 41]. Hereafter, we consider a dictionary made up of discrete

real-valued curvelets [11]; curvelets are known to yield sparse representations of piecewise smooth functions. The choice of a curvelet dictionary is also motivated by the existence of fast algorithms for the computation of the product between \mathbf{D} and some vector (see [10]): this transform is based on a fast Fourier transform and its complexity⁹ scales as $\mathcal{O}(n \log n)$.

Matrix \mathbf{H} appearing in the cost-to-move function in (4.18) is chosen to be a Haar wavelet basis.¹⁰ In practice, we did not observe a significant difference in our results by using other types of wavelets; we thus essentially consider Haar wavelets for simplicity purposes.

To complete our discussion, let us elaborate on the choice of \mathbf{G} and \mathcal{R} , characterizing the regularization imposed on the displacement field \mathbf{d}_t . In our simulations, we wish to enforce either a global or a piecewise regularity of the motion. We proceed as follows. The spatial derivatives of the motion are approximated by a “finite difference” scheme: each finite difference corresponds to a particular element of the matrix-vector product $\mathbf{G}^* \mathbf{d}_t$ (matrix \mathbf{G} thus contains “ ± 1 ” elements located at proper positions). The regularity of the motion field is then enforced by constraining the function $\mathcal{R}(\mathbf{G}^* \mathbf{d}_t)$ to be small. In our experimentations, we choose \mathcal{R} to be defined as in (3.5) with a weighting vector \mathbf{w} as in [55]. Further details are provided in Appendix C.1.

Besides, we notice that, although we have presented our SR procedure in the case of a monochannel image-sequence observation in section 4, its extension to a multichannel setting (e.g., when 3-channel color images are available) is straightforward and will be considered in our simulations.

We now specify the choice of the algorithm parameters. As exposed earlier, we rely on the recursion (4.17)–(4.19) described in section 4.3 to search for a critical point of the cost function in (4.1) with $p = 1$. Each step of the recursion (4.17)–(4.19) is solved via an ADMM procedure. Details on the ADMM steps are given in Appendix C.2. The ADMM solvers involve minimizations by a gradient descent procedure. In our implementation, we choose a quasi-Newton descent method adapted to our high-dimensional problem, namely, a limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) procedure with a line-search routine based on the strong Wolfe conditions [37]. We stop the ADMM recursions after 20 iterations and the global 2-step recursions after 20 iterations, since we observed no significant improvements of the results for a larger number of iterations.

The superresolved images \mathbf{x}_t ’s are initialized by Lanczos interpolation of their LR counterparts. Motion fields are initialized with an upscaled optic-flow estimate obtained by applying algorithm [55] on the LR observations. To perform a fair comparison with the multiframe SR algorithm of Mitzel et al. [35] described in the next section, we also ran our algorithm with an initial motion field computed with the optic-flow algorithm [59]. For both initializations, the upscaling from the LR optic-flow estimate to the HR motion field is done with a Lanczos interpolation. The values of the other parameters of our algorithm are given in Table 1. These parameters have been tuned experimentally to lead to a reasonable trade-off between visual inspection and error measurements for the data-set benchmark presented in section 5.3.

⁹As mentioned earlier, a linear complexity can be preserved by using, for example, a wavelet basis instead of a curvelet frame.

¹⁰We note that evaluation of products \mathbf{H} or \mathbf{H}^* only requires a linear complexity since they can be implemented by fast wavelet transforms [33, Chapter 7].

Table 1

Algorithm parameter setting. Parameters α_1 , α_2 , and α_3 appear in the cost function (4.1). Parameters γ and $\alpha^{(0)}$ specify the 2 steps (4.17) and (4.19). Parameters ρ_1 , ρ_2 , ρ_3 , and ρ are auxiliary factors used in the ADMM recursions.

α_1	ρ_1	α_2	ρ_2	α_3	ρ_3	γ	ρ	$\alpha^{(0)}$
5e-1	1e2	8e3	1e1	1e1	1e-2	1e0	1e0	2e2

5.2. Algorithm benchmark. The assessment of the proposed algorithm relies on a comparison with a benchmark of three state-of-the-art methods:

- the single-frame SR algorithm of Peleg and Elad, 2014 [41],
- the kernel-regression SR algorithm of Takeda et al., 2009 [47, 46],
- the multi-frame SR algorithm of Mitzel et al., 2009 [35].

These algorithms are adapted to the SR of videos exhibiting nonhomogeneous displacements. Moreover, each of these three algorithms is a state-of-the-art method representing a class of SR algorithms. The algorithm of Peleg et al., 2014, implements a single-frame SR method based on a statistical learning procedure with sparse representations; the algorithm by Takeda et al., 2009, is an SR method based on a “multidimensional kernel regression” fitting the LR observations; the algorithm of Mitzel et al., 2009, implements a multiframe SR method using a quadratic relaxation scheme for high-accuracy optic-flow estimation [59]. Finally, we also compare the performance obtained with the proposed method with two standard spatial interpolation techniques, namely,

- the basic *nearest neighbor* upscaling (block interpolation),
- *Lanczos* interpolation [50].

Note that in order to treat color image sequences, algorithms only supporting gray-level images are run independently on the three spectral bands.

5.3. Data-set benchmark. We evaluate the performance of the algorithms using a benchmark of three image sequences:

- A synthetic sequence from the *MPI Sintel* data set [9]. This recent data set, which is derived from the open source three-dimensional animated short film, was originally created for the evaluation of optical flows. The synthesized image sequences are realistic and particularly challenging: on the one hand, displacement fields are characterized by large amplitudes, discontinuities, blur or defocus effects; on the other hand, the image sequence presents many occlusions, specular reflections, or atmospheric effects. In our simulations, we focus on a region of interest of 436×512 pixels and on the first 8 images. The first and last images of the “*bandage*” data-set sequence are displayed in Figure 1. In the following, we will refer to this sequence as data set #1.
- A real sample of the standard “*foreman*” video.¹¹ In our simulations, we focus on a region of interest of 256×256 pixels and on the 10 first images. The first and last images of this data set are displayed in Figure 2. In the following, we will refer to this sequence as data set #2.

¹¹Image sequences are part of the *Derf Collection*, which can be downloaded at <https://media.xiph.org/video/derf> (2015).



Figure 1. Data set #1: first and last frames of the “bandage” sequence.



Figure 2. Data set #2: first and last frames of the “foreman” sequence.



Figure 3. Data set #3: first and last frames of the “football” sequence.

- A real sample of the challenging “football” video¹¹, which exhibits nonhomogeneous and large displacements, as well as multiple occlusions. In our simulations, we focus on a region of interest of 256×256 pixels and on the 10 first images. The first and last images of this data set are displayed in Figure 3. In the following, we will refer to this sequence as data set #3.

The images of these sequences are composed of three spectral bands, each one is coded in 8 bits. We create the LR images by applying the function \mathcal{H} on these sequences. This function first filters the discrete signal by a Gaussian kernel of standard deviation equal to 1.12 and then down samples the result by a factor of 2; see [8].

5.4. Evaluation procedure. The performance of the algorithms is assessed in terms of reconstruction of the superresolved image and estimation of the motion field. We describe the figures of merit used in our assessments hereafter. Let $\{\hat{\mathbf{x}}_t\}_{t=0}^T$ (resp., $\{\hat{\mathbf{d}}_t\}_{t=1}^T$) denote the estimated image sequence (resp., displacements) and $\{\mathbf{x}_t^{true}\}_{t=0}^T$ (resp., $\{\mathbf{d}_t^{true}\}_{t=1}^T$) the corresponding ground truth. Standard criteria [36] to measure the image sequence reconstruction accuracy are the peak signal to noise ratio (PSNR) at time t ,

$$\text{PSNR}(t) = 20 \log_{10} \frac{n \|\mathbf{x}_t^{true}\|_{\infty}}{\|\mathbf{x}_t^{true} - \hat{\mathbf{x}}_t\|_2},$$

and the correlation coefficient (CC) at time t ,

$$\text{CC}(t) = \frac{(\mathbf{x}_t^{true} - \mu_{\mathbf{x}_t^{true}})^* (\hat{\mathbf{x}}_t - \mu_{\hat{\mathbf{x}}_t})}{\|\mathbf{x}_t^{true} - \mu_{\mathbf{x}_t^{true}}\|_2 \|\hat{\mathbf{x}}_t - \mu_{\hat{\mathbf{x}}_t}\|_2},$$

where we have denoted the arithmetic mean of vectors $\hat{\mathbf{x}}_t$ and \mathbf{x}_t^{true} by $\mu_{\hat{\mathbf{x}}_t}$ and $\mu_{\mathbf{x}_t^{true}}$. We evaluate the accuracy of the estimated motion fields with the time-averaged mean end point error (MEPE),

$$\text{MEPE} = \frac{1}{nT} \sum_{t=1}^T \|\mathbf{d}_t^{true} - \hat{\mathbf{d}}_t\|_1,$$

and the time-averaged mean Barron angular error (MBAE) in degrees [5],

$$\text{MBAE} = \frac{1}{nT} \sum_{t=1}^T \sum_{s=1}^n \arccos \left(\frac{1 + \mathbf{d}_t^{true}(s) \hat{\mathbf{d}}_t(s) + \mathbf{d}_t^{true}(s+n) \hat{\mathbf{d}}_t(s+n)}{\sqrt{(1 + \hat{\mathbf{d}}_t(s)^2 + \hat{\mathbf{d}}_t(s+n)^2) (1 + \mathbf{d}_t^{true}(s)^2 + \mathbf{d}_t^{true}(s+n)^2)}} \right),$$

where we have adopted the convention that the two n -dimensional components of motion have been sorted one after the other in vectors $\hat{\mathbf{d}}_t$ and \mathbf{d}_t^{true} .

In order to compare the different algorithms (algorithm [47] does not support large images and excludes pixels at the image border), the criteria PSNR and CC are evaluated on a spatial window of size 240×240 cropped in the image sequences.

5.5. Results and discussion. Table 2 presents the accuracy of the different algorithms in terms of PSNR and CC. We evaluated these criteria at $t = 5$ for data set #1 and at $t = 7$ for data sets #2 and #3. We first note that our SR method yields better figures of merit than the other methods for the different data sets of the benchmark. It improves slightly the CC and substantially the PSNR (more than a unit) for each data-set configuration. Second, the estimates released by the proposed approach seem to achieve a good quality level irrespective

Table 2

Accuracy of super-resolved image estimates in terms of PSNR and CC at time t .

	PSNR(t)			CC(t)		
	Set # 1	Set # 2	Set # 3	Set # 1	Set # 2	Set # 3
	$t=5$	$t=7$	$t=7$	$t=5$	$t=7$	$t=7$
Nearest neighbor	27.236	22.639	22.656	0.9771	0.9648	0.9556
Lanczos interpolation [50]	27.845	24.571	23.040	0.9800	0.9775	0.9579
Single-frame SR [41]	28.359	32.518	23.971	0.9815	0.9949	0.9684
Kernel-regression SR [47, 46]	28.944	33.275	23.394	0.9838	0.9957	0.9643
Multiframe SR [35]	29.935	32.295	21.948	0.9844	0.9939	0.9491
Proposed (optic-flow init. [59])	30.634	35.027	25.007	0.9868	0.9969	0.9750
Proposed (optic-flow init. [55])	30.790	34.305	25.302	0.9872	0.9963	0.9767

of the considered data set: on the contrary, the multiframe SR algorithm [35] performs fairly well on data set #1 but its performance collapses on data set #3; the kernel-regression SR algorithm [47, 46] obtains good results for data set #2 while yielding only a slight increase of the accuracy with respect to a Lanczos interpolation on data set #3; the single-frame SR algorithm [41] has a good behavior for data set #3 but is less competitive for data set #1. Third, the performance of our algorithm seems to be comparable for different motion initializations, in particular for initial motion fields obtained from the optic-flow algorithms of [55] or [59].

The improvement brought by the proposed method can also be seen by a visual inspection of the reconstructed images in Figures 4, 5, 6, and 7. We can first underline the enhancement provided by the inclusion of some motion information in the SR reconstruction process by comparing the estimates released by the single-frame and the multiframe/sequential algorithms. In Figures 4 and 6, one can notice that the estimated contours and the texture are oversmoothed if no motion information is included. This is, for example, visible by inspecting the fuzzy girl’s eyebrow or the smoothed scales of the little dragon in Figure 4, distinguishing the tongue of the foreman or analyzing the texture of the grass field of the football game in Figure 6. In comparison, our algorithm enhances the reconstruction accuracy of these details as is visible in Figures 5 and 7. The drawback of including motion is that, as can be noticed for the little dragon, errors in motion discontinuity estimation may induce imprecision on the contours and lead to some undesirable oscillations.

Although not as accurate as the proposed method, we note the good performance of the single-frame SR algorithm proposed in [41]. Clearly, it is competitive with other state-of-the-art approaches exploiting motion information. This is probably due to the relevance of the sparse prior employed by the single-frame SR algorithm [41]. This is particularly striking when the motion in the video is too difficult to exploit by the multiframe or kernel-regression SR algorithms, as shown for the challenging football sequence in Figures 6 and 7. Let us also mention that results obtained with a kernel-regression SR strategy reveal a slight enhancement in comparison to standard spatial interpolation techniques, which is probably induced by the implicit introduction of the motion information via the modeling of the local spatiotemporal structures of the sequence.

Our experiments also emphasize several examples where a sequential SR setup can solve some reconstruction ambiguities which can be difficult to treat in a multiframe framework. In Figures 5 and 7, some erroneous reconstructions, which do not appear in the proposed



Figure 4. *Single-image SR estimates for data set #1. Details of the SR images obtained with nearest neighbor strategy (first row), Lanczos interpolation (second row), and the learning algorithm proposed in [41] (third row).*

method, can be noticed in the multiframe estimates: for example, artifacts in the girl's eye in Figure 5, deformations of the foreman's tongue and the fuzziness of the stripes of the football player's trousers in Figure 7. Indeed, matching all the images of the sequence with a reference frame is often a more difficult task than estimating motions between consecutive frames. In the former situation, motion estimation has to deal with large displacements between distant frames whereas, in the latter setup, the problem simplifies to the estimation of a succession of small displacements. In other words, an SR multiframe setup will try to match images of the sequence which could apparently seem independent, with the potential drawback of estimating erroneous structures. On the other hand, an SR sequential setup propagates information through consecutive frames and may better succeed in modeling the overall dependences in the image sequence. One could nevertheless argue that the estimation of interframe motions could also lead to error propagation if the motion estimates are inaccurate. This is not what we observed in our simulations: motion errors are usually absorbed by the error terms ϵ_t

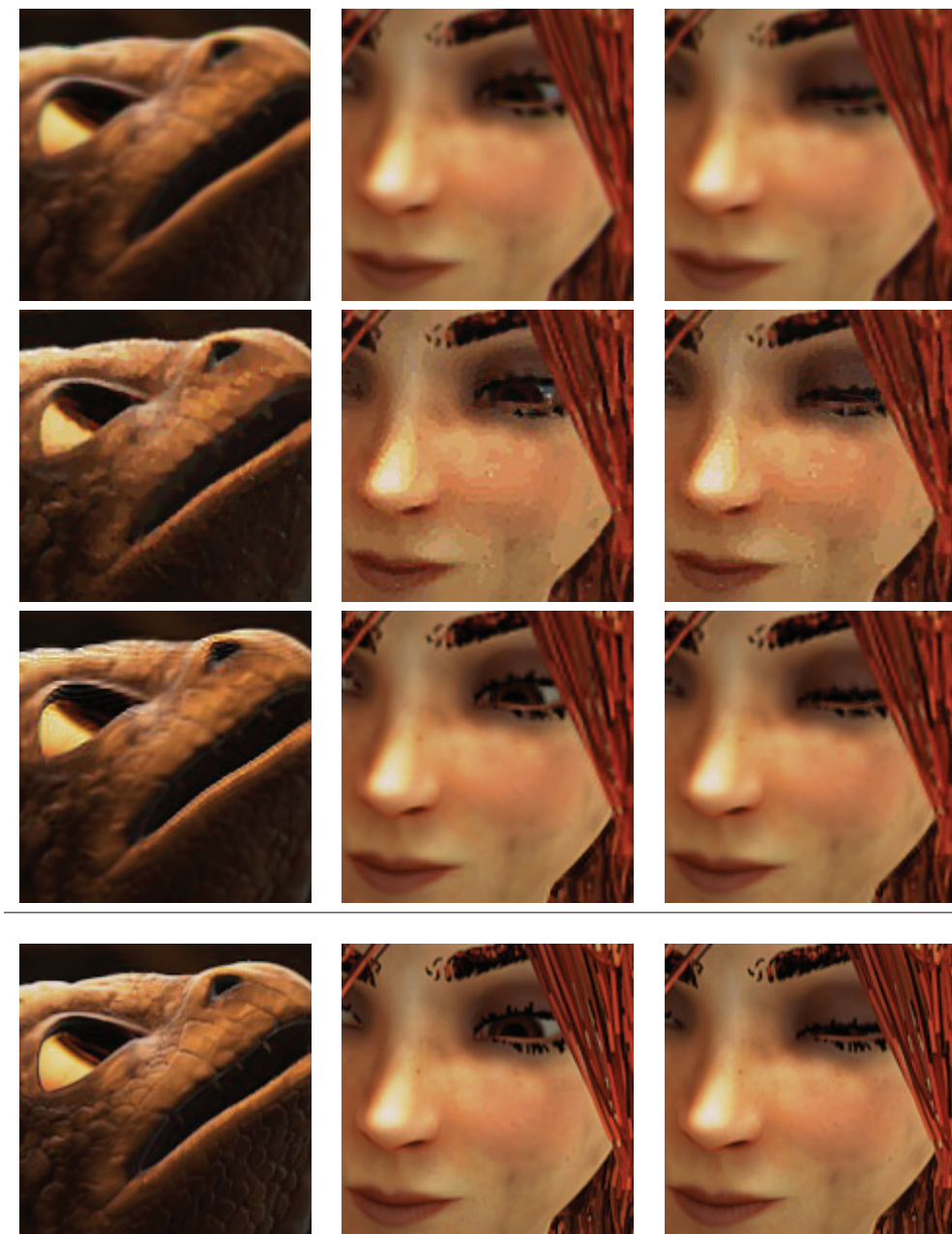


Figure 5. Multiframe and sequential SR estimates for data set #1. Details of the SR images obtained with the multiframe algorithms of [47, 46] (first row) or [35] (second row), and with the proposed sequential algorithm (third row) in comparison to ground truth (fourth row). Initialization of our algorithm relies on the optic-flow method [55].

(which increase in the region where the motion is badly estimated). This is illustrated in Figure 8: we observe that ϵ_t may be large on the contours of the characters (where the quality of the motion estimation is typically low) but the PSNR is nevertheless stable across the

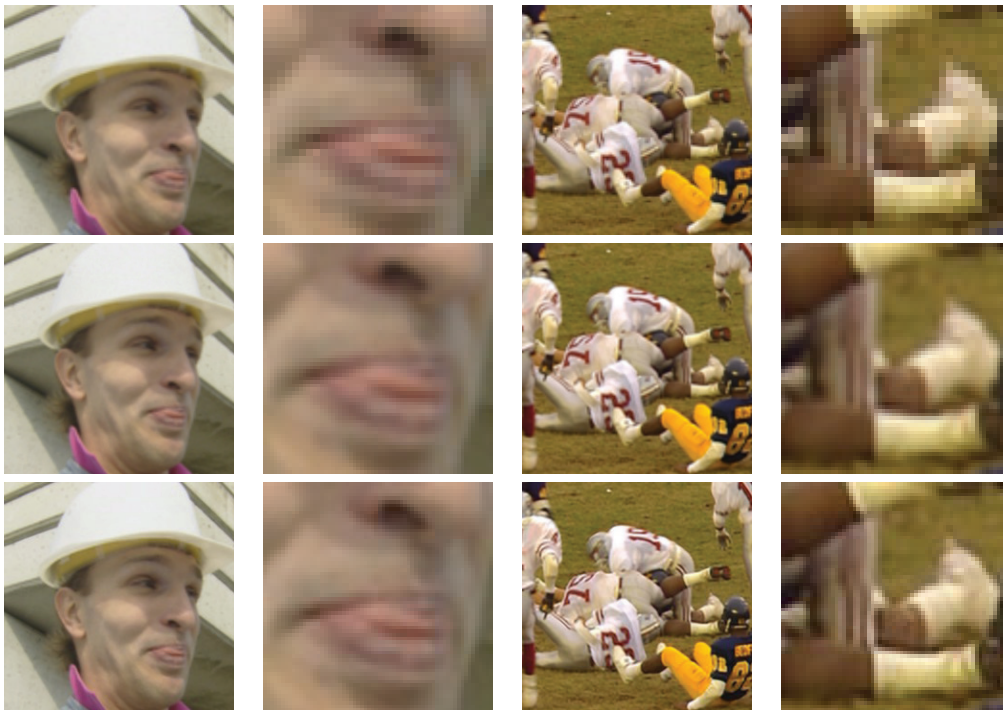


Figure 6. Single-image SR estimates for data sets #2 and #3. SR images and details obtained with nearest neighbor strategy (first row), Lanczos interpolation (second row), and the learning algorithm of [41] (third row).

reconstructed image sequence. Therefore, as observed from our simulations, a sequential SR approach is usually better conditioned to deal with videos such as data sets #1 and #3, which exhibit large displacements and/or occlusions.

Finally, let us notice that there is a positive interaction between the estimation of the motion fields and the HR images: intuitively, it is clear that a good estimation of the HR image sequence will improve the quality of the estimated motion fields; similarly, a good estimation of the superresolved motion fields will enhance the accuracy of the estimated image sequence. Although this positive interaction is difficult to ensure from a theoretical side, we have often observed it in practice. We illustrate in Table 3 and Figure 9 the benefit of refining the motion estimation through our iterative procedure for the synthetic data set #1, independently of the initial motion estimate. In Table 3, we can notice a slight gain in terms of MBE and MEPE in comparison to a direct estimation of the motion from the LR observations with the methods presented in [55] or [59] (which serve as initializations for our algorithm; see section 5.1). More interestingly, we note in Figure 9 that the motion field released by the proposed approach exhibits sharper discontinuities than those output by [55] or [59].

6. Conclusion. We have presented a new methodology to solve video SR problems, i.e., to reconstruct an HR image sequence from LR observations. The HR sequence is entirely described by a parametric nonlinear sequential model, which connects the different images of

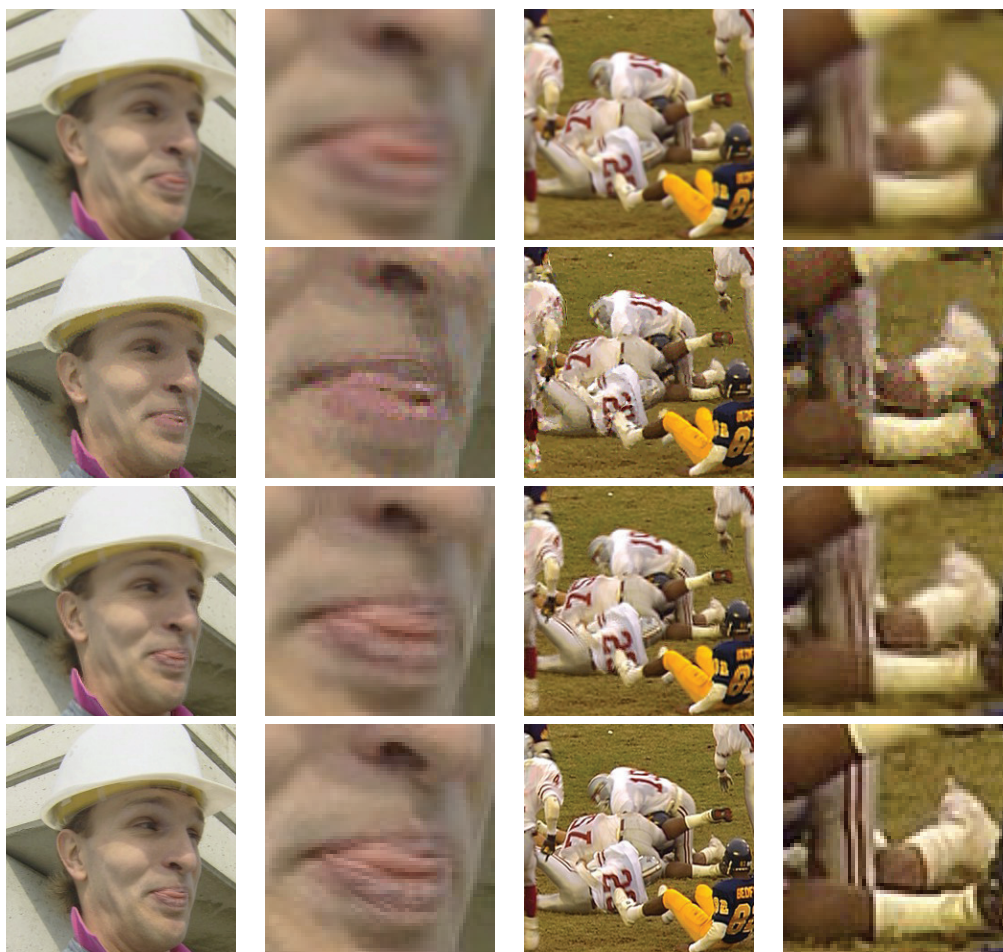


Figure 7. Multiframe or sequential SR estimates for data sets #2 and #3. SR images and details obtained with the multiframe algorithms of [47, 46] (first row) or [35] (second row), and with the proposed sequential algorithm (third row) in comparison to ground truth (fourth row). Initialization of our algorithm relies on the optic-flow method [55].

the sequence. It is parametrized by a final condition, a sequence of nonglobal displacement fields and a sequence of additive noises. In order to compensate for the ill-posedness of the video SR problem, we considered priors enforcing some forms of sparsity on the unknown parameters of the system. The joint estimation of the final condition, the displacement, and the noise sequences was expressed as a constrained minimization problem which, in the general case, is high dimensional, nondifferentiable and nonconvex. We provided elementary building blocks to tackle each of these difficulties, and, by gathering them, designed a convergent optimization algorithm enjoying a complexity (per iteration) linear in the problem dimensions. Our numerical simulations on several video benchmarks show that the proposed SR method is competitive with state of the art. In particular, the gain appears to be particularly important for videos involving complex motions with large amplitudes and occlusions.

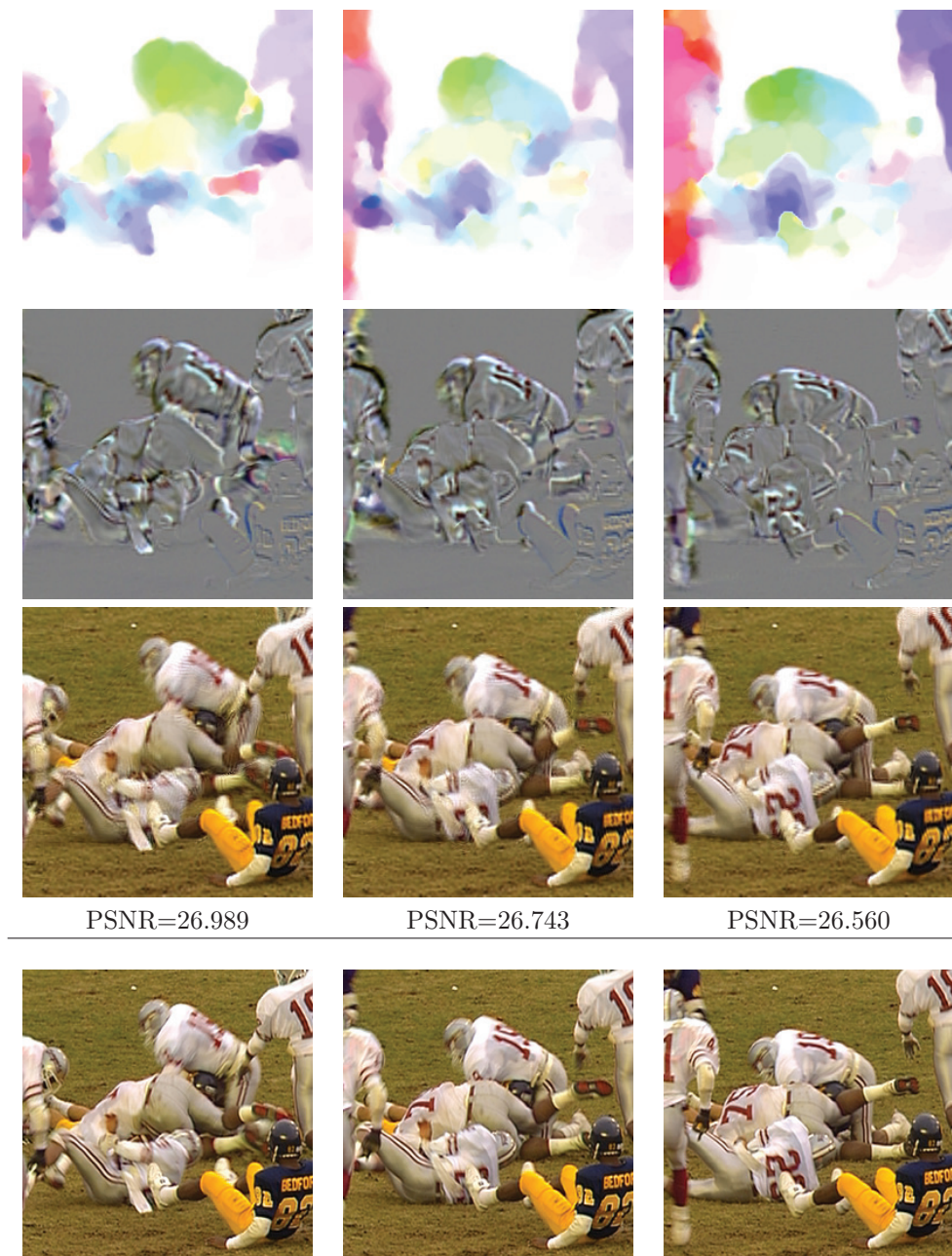


Figure 8. Reconstruction of three superresolved images \mathbf{x}_t (third row), optic-flow fields \mathbf{d}_t (first row), and warping errors ϵ_t (second row) for data set #3 corresponding to $t = 3$ (left), $t = 5$ (middle), and $t = 7$ (right). True images (fourth row) and associated PSNR (computed without quantification of the estimates and including the image borders) are displayed below. Initialization of our algorithm relies on the optic-flow method [55].

Table 3

Accuracy of low-resolved or superresolved optic-flow estimates in terms of MEPE and MBAE, with respect to the motion initialization algorithm.

	LR estimate		SR estimate	
	MEPE	MBAE	MEPE	MBAE
Zach, Pock, and Bischof 2007 [59].	1.319	24.988	1.302	24.975
Xu, Jia, and Matsushita, 2012 [55].	1.342	25.592	1.320	25.545

Appendix A. Proof of (4.5)–(4.7). The proof of this specific *backward* optimal control solution follows the sketch of the demonstration for the more standard *forward* problem presented in [6]. We will focus on the following optimization problem

$$(A.1) \quad \arg \min_{(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})} \mathcal{T}(\mathbf{x} = \mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}), \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}),$$

where \mathcal{T} denotes some objective function to be defined below. We recall that, given $(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$, the function $\mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ determines a unique vector $\mathbf{x} = \mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ satisfying the constraints in (4.1); see section 4.1. In this appendix, we will use the following shorthand notation for the constraints in (4.1):

$$(A.2) \quad \begin{cases} \mathbf{x}_t = \mathcal{F}_t(\mathbf{x}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}), & 0 \leq t \leq T-1, \\ \mathbf{x}_T = \mathbf{D}\mathbf{c} \end{cases}$$

with $\mathcal{F}_t(\mathbf{x}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}) \triangleq \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}$. We will also alleviate the notation for the constraint $\mathbf{x} = \mathcal{Q}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ by denoting this vector simply by \mathbf{x} . Therefore, \mathbf{x} should be understood as a function of $\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}$ and no longer as an independent variable.

The proof of (4.5)–(4.7) is made of two different parts, in which we study different instances of optimization problem (A.1). In a first step, we will consider an objective function only depending on the initial state:¹²

$$(A.3) \quad \mathcal{T}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) \triangleq \mathcal{G}_0(\mathbf{x}_0).$$

Then, in a second step, we will come back to the more general problem (4.2), i.e., an optimization problem where the objective function \mathcal{T} will match the cost function, \mathcal{J} , given in (4.3):

$$\mathcal{T}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}) \triangleq \mathcal{G}_0(\mathbf{x}_0) + \sum_{t=1}^{T-1} \mathcal{G}_t(\mathbf{x}_t, \boldsymbol{\epsilon}_t, \mathbf{d}_t) + \mathcal{G}_T(\mathbf{x}_T, \boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}) \triangleq \mathcal{J}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}).$$

First part of the proof. We begin by considering problem (A.1) with the objective function (A.3). By the chain rule of derivation applied to (A.2) at some point in the set

$$(A.4) \quad \left\{ (\mathbf{x}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') \left| \begin{array}{l} \mathbf{x}'_t = \mathcal{F}_t(\mathbf{x}'_{t+1}, \boldsymbol{\epsilon}'_{t+1}, \mathbf{d}'_{t+1}), \quad 0 \leq t \leq T-1 \\ \mathbf{x}'_T = \mathbf{D}\mathbf{c}' \end{array} \right. \right\},$$

¹²As mentioned previously, \mathbf{x}_0 must be understood as a function of $\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c}$.

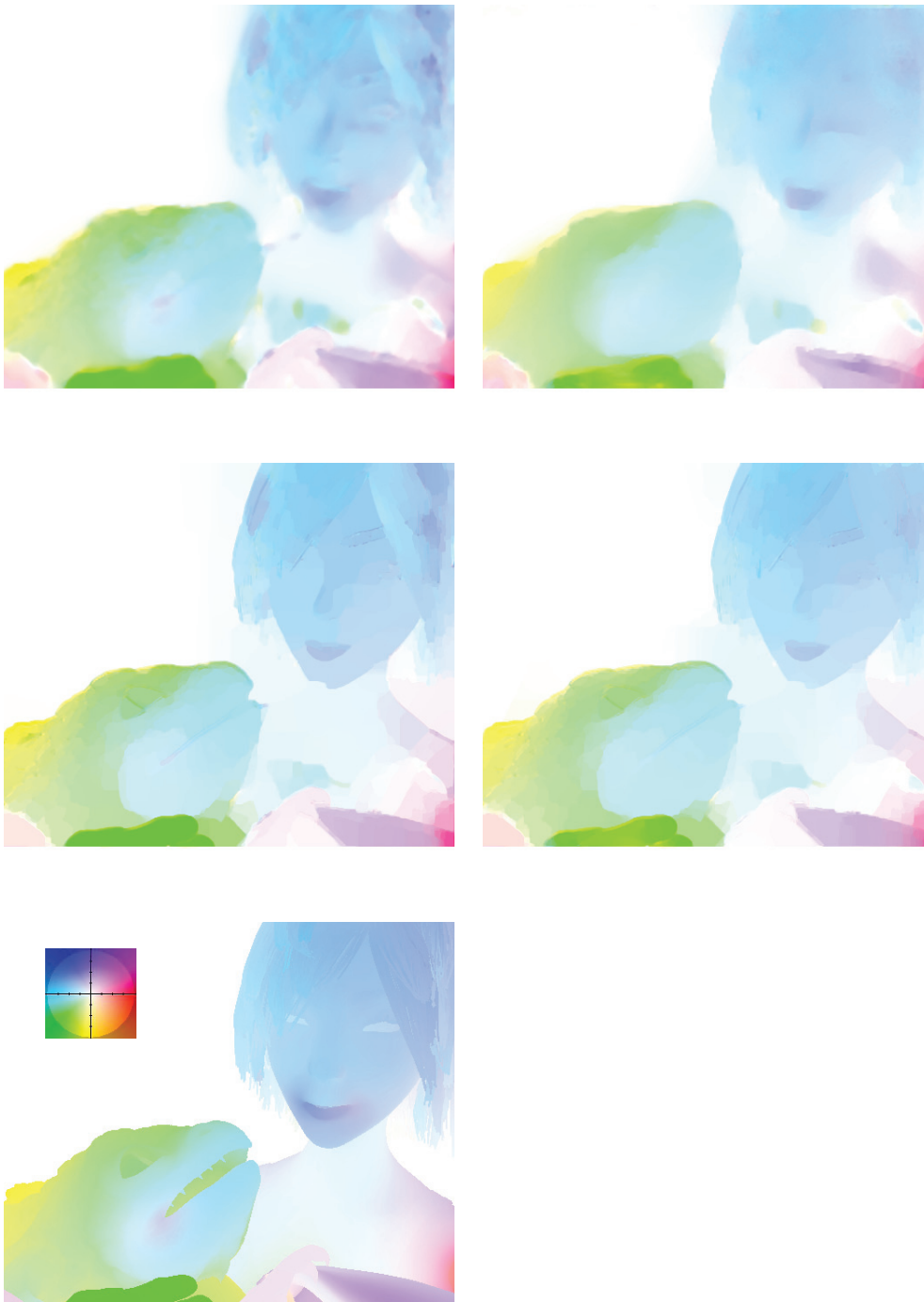


Figure 9. Optic-flow SR estimates. Motion field estimated from low-resolved images of data set #1 at initial time. Top: estimates for state-of-the-art algorithms [55] (left) and [59] (right). Middle: estimates with the proposed SR algorithm initialized with [55] (left) or [59] (right). Bottom: ground truth and associated colormap.

we can decompose the gradients into the products

$$(A.5) \quad \begin{cases} \nabla_{\epsilon_t} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \nabla_{\epsilon_t} \mathcal{F}_{t-1}^*(\mathbf{x}'_t, \epsilon'_t, \mathbf{d}'_t) \nabla_{\mathbf{x}_{t-1}} \mathcal{F}_{t-2}^* \cdots \nabla_{\mathbf{x}_2} \mathcal{F}_1^* \nabla_{\mathbf{x}_1} \mathcal{F}_0^* \nabla_{\mathbf{x}_0} \mathcal{G}_0(\mathbf{x}'_0), \\ \nabla_{\mathbf{d}_t} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \nabla_{\mathbf{d}_t} \mathcal{F}_{t-1}^*(\mathbf{x}'_t, \epsilon'_t, \mathbf{d}'_t) \nabla_{\mathbf{x}_{t-1}} \mathcal{F}_{t-2}^* \cdots \nabla_{\mathbf{x}_2} \mathcal{F}_1^* \nabla_{\mathbf{x}_1} \mathcal{F}_0^* \nabla_{\mathbf{x}_0} \mathcal{G}_0(\mathbf{x}'_0), \\ \nabla_{\mathbf{c}} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \mathbf{D}^* \nabla_{\mathbf{x}_T} \mathcal{F}_{T-1}^* \cdots \nabla_{\mathbf{x}_2} \mathcal{F}_1^* \nabla_{\mathbf{x}_1} \mathcal{F}_0^* \nabla_{\mathbf{x}_0} \mathcal{G}_0(\mathbf{x}'_0), \end{cases}$$

where we recall that $\nabla_{\mathbf{x}_t} \mathcal{F}_{t-1}$ denotes the Jacobian matrix of \mathcal{F}_{t-1} with respect to function \mathbf{x}_t evaluated at $(\mathbf{x}'_t, \epsilon'_t, \mathbf{d}'_t)$ and $\nabla_{\mathbf{x}_t} \mathcal{F}_{t-1}^*$ its transpose. We can rewrite gradients in (A.5) in order to exhibit their recursive structures. By defining the forward recursion

$$(A.6) \quad \begin{cases} \zeta_0 = \nabla_{\mathbf{x}_0} \mathcal{G}_0(\mathbf{x}'_0), \\ \zeta_t = \nabla_{\mathbf{x}_t} \mathcal{F}_{t-1}^* \zeta_{t-1}, \quad 1 \leq t \leq T, \end{cases}$$

we obtain the following rewriting:

$$(A.7) \quad \begin{cases} \nabla_{\epsilon_t} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \nabla_{\epsilon_t} \mathcal{F}_{t-1}^*(\mathbf{x}'_t, \epsilon'_t, \mathbf{d}'_t) \zeta_{t-1}, \quad 1 \leq t \leq T, \\ \nabla_{\mathbf{d}_t} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \nabla_{\mathbf{d}_t} \mathcal{F}_{t-1}^*(\mathbf{x}'_t, \epsilon'_t, \mathbf{d}'_t) \zeta_{t-1}, \quad 1 \leq t \leq T, \\ \nabla_{\mathbf{c}} \mathcal{T}(\mathbf{x}', \epsilon', \mathbf{d}', \mathbf{c}') = \mathbf{D}^* \zeta_T. \end{cases}$$

Second part of the proof. We now consider problem (A.1) with objective function (A). By making a change of variables, we want to obtain a rewriting of function (A) with a structure analogous to (A.3), so that the gradients are given by a recursion of the form (A.6)–(A.7). In other words, by making some change of variables we intend to rewrite the sum of functions in (A) as a unique function depending solely on an “initial state.” In order to do so, let us define variables κ_i ’s recursively as follows:

$$\begin{cases} \kappa_T = 0, \\ \kappa_{T-1}(\mathbf{x}_T, \epsilon_T, \mathbf{d}_T, \mathbf{c}) = \kappa_T + \mathcal{G}_T(\mathbf{x}_T, \epsilon_T, \mathbf{d}_T, \mathbf{c}), \\ \kappa_{t-1}(\mathbf{x}_t, \epsilon_t, \mathbf{d}_t, \mathbf{c}) = \kappa_t + \mathcal{G}_t(\mathbf{x}_t, \epsilon_t, \mathbf{d}_t, \mathbf{c}), \quad T-1 \geq t \geq 1. \end{cases}$$

We then obtain that

$$\kappa_0(\mathbf{x}, \epsilon, \mathbf{d}, \mathbf{c}) = \sum_{t=1}^{T-1} \mathcal{G}_t(\mathbf{x}_t, \epsilon_t, \mathbf{d}_t) + \mathcal{G}_T(\mathbf{x}_T, \epsilon_T, \mathbf{d}_T, \mathbf{c}),$$

and the objective function \mathcal{T} given in (A) can be rewritten as

$$(A.8) \quad \mathcal{T}(\mathbf{x}, \epsilon, \mathbf{d}, \mathbf{c}) = \kappa_0(\mathbf{x}, \epsilon, \mathbf{d}, \mathbf{c}) + \mathcal{G}_0(\mathbf{x}_0).$$

Considering the following change of variables $\tilde{\mathbf{x}}_t \triangleq \begin{pmatrix} \mathbf{x}_t \\ \kappa_t \end{pmatrix}$, we then have that the right-hand side of (A.8) can be rewritten as a function of $\tilde{\mathbf{x}}_0$ only. In what follows, we will use the following specific notation to emphasize this fact:

$$(A.9) \quad \tilde{\mathcal{G}}_0(\tilde{\mathbf{x}}_0) = \kappa_0 + \mathcal{G}_0(\mathbf{x}_0).$$

Moreover, it is easy to see that the functions $\tilde{\mathbf{x}}_t$'s satisfy the following backward recursion:

$$(A.10) \quad \begin{cases} \tilde{\mathbf{x}}_t = \tilde{\mathcal{F}}_t(\tilde{\mathbf{x}}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}), & T-1 \geq t \geq 0, \\ \tilde{\mathbf{x}}_T = \tilde{\mathcal{F}}_T(\boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}), \end{cases}$$

where

$$\begin{aligned} \tilde{\mathcal{F}}_t(\tilde{\mathbf{x}}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}) &= \begin{pmatrix} \mathcal{F}_t(\mathbf{x}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}) \\ \boldsymbol{\kappa}_{t+1} + \mathcal{G}_{t+1}(\mathbf{x}_{t+1}, \boldsymbol{\epsilon}_{t+1}, \mathbf{d}_{t+1}) \end{pmatrix}, \\ \tilde{\mathcal{F}}_T(\boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}) &= \begin{pmatrix} \mathbf{Dc} \\ \mathcal{G}_T(\mathbf{Dc}, \boldsymbol{\epsilon}_T, \mathbf{d}_T, \mathbf{c}) \end{pmatrix}. \end{aligned}$$

We remark that the cost function (A.9), recursion (A.10), and the set

$$(A.11) \quad \left\{ (\tilde{\mathbf{x}}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') \left| \begin{array}{l} \tilde{\mathbf{x}}'_t = \tilde{\mathcal{F}}_t(\tilde{\mathbf{x}}'_{t+1}, \boldsymbol{\epsilon}'_{t+1}, \mathbf{d}'_{t+1}), \quad 0 \leq t \leq T-1 \\ \tilde{\mathbf{x}}'_T = \tilde{\mathcal{F}}_T(\boldsymbol{\epsilon}'_T, \mathbf{d}'_T, \mathbf{c}') \end{array} \right. \right\},$$

have, respectively, the same structure as (A.3), (A.2), and (A.4). We can then apply the result obtained previously and get the gradients of \mathcal{T} using the same reasoning as the one made to derive (A.6)–(A.7). More specifically, let $(\tilde{\mathbf{x}}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}')$ be some point in (A.11), and let $\tilde{\boldsymbol{\zeta}}_t$ be an adjoint variable verifying

$$(A.12) \quad \begin{cases} \tilde{\boldsymbol{\zeta}}_0 = \nabla_{\tilde{\mathbf{x}}_0} \tilde{\mathcal{G}}_0(\tilde{\mathbf{x}}'_0), \\ \tilde{\boldsymbol{\zeta}}_t = \nabla_{\tilde{\mathbf{x}}_t} \tilde{\mathcal{F}}_{t-1}^* \tilde{\boldsymbol{\zeta}}_{t-1}, \quad 1 \leq t \leq T, \end{cases}$$

where the Jacobian matrix of $\tilde{\mathcal{F}}_{t-1}$ evaluated at some point $(\tilde{\mathbf{x}}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t)$ is denoted $\nabla_{\tilde{\mathbf{x}}_t} \tilde{\mathcal{F}}_{t-1}$. Using (A.7), we obtain the following expressions:

$$\begin{cases} \nabla_{\mathbf{d}_t} \mathcal{T}(\tilde{\mathbf{x}}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \nabla_{\mathbf{d}_t} \tilde{\mathcal{F}}_{t-1}^*(\tilde{\mathbf{x}}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t) \tilde{\boldsymbol{\zeta}}_{t-1}, & 1 \leq t \leq T, \\ \nabla_{\boldsymbol{\epsilon}_t} \mathcal{T}(\tilde{\mathbf{x}}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \nabla_{\boldsymbol{\epsilon}_t} \tilde{\mathcal{F}}_{t-1}^*(\tilde{\mathbf{x}}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t) \tilde{\boldsymbol{\zeta}}_{t-1}, & 1 \leq t \leq T, \\ \nabla_{\mathbf{c}} \mathcal{T}(\tilde{\mathbf{x}}', \boldsymbol{\epsilon}', \mathbf{d}', \mathbf{c}') = \nabla_{\mathbf{c}} \tilde{\mathcal{F}}_T^*(\boldsymbol{\epsilon}'_T, \mathbf{d}'_T, \mathbf{c}') \tilde{\boldsymbol{\zeta}}_T. \end{cases}$$

To finalize the proof, we reexpress recursion (A.12) by developing it with respect to the two different components of the adjoint variable $\tilde{\boldsymbol{\zeta}}_t \triangleq \begin{pmatrix} \boldsymbol{\zeta}_t \\ \omega_t \end{pmatrix}$, where the $\boldsymbol{\zeta}_t$'s have the dimension of the \mathbf{x}_t 's and ω_t 's are scalars. Particularizing the first equation in (A.12) by taking (A.9) into account, we obtain

$$(A.13) \quad \begin{pmatrix} \boldsymbol{\zeta}_0 \\ \omega_0 \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}_0} \tilde{\mathcal{G}}_0(\tilde{\mathbf{x}}'_0) \\ \nabla_{\boldsymbol{\kappa}_0} \tilde{\mathcal{G}}_0(\tilde{\mathbf{x}}'_0) \end{pmatrix} = \begin{pmatrix} \nabla \mathcal{G}_0(\mathbf{x}'_0) \\ 1 \end{pmatrix}.$$

Moreover, using the definition of $\tilde{\mathcal{F}}_t$, the second equation in (A.12) leads to

$$(A.14) \quad \begin{pmatrix} \boldsymbol{\zeta}_t \\ \omega_t \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}_t} \mathcal{F}_{t-1}^* & \nabla_{\mathbf{x}_t} \mathcal{G}_t(\mathbf{x}'_t, \boldsymbol{\epsilon}'_t, \mathbf{d}'_t) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_{t-1} \\ \omega_{t-1} \end{pmatrix}, \quad 1 \leq t \leq T-1,$$

$$(A.15) \quad \begin{pmatrix} \boldsymbol{\zeta}_T \\ \omega_T \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}_T} \mathcal{F}_{T-1}^* & \nabla_{\mathbf{x}_T} \mathcal{G}_T(\mathbf{x}'_T, \boldsymbol{\epsilon}'_T, \mathbf{d}'_T, \mathbf{c}') \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_{T-1} \\ \omega_{T-1} \end{pmatrix}.$$

Equations (A.13)–(A.15) imply that $\omega_t = 1 \forall t$; moreover, the recursion in ζ_t is equivalent to (4.7).

Appendix B. The ADMM. The ADMM focuses on the following type of optimization problems:

$$(B.1) \quad \min_{z_1 \in \Xi_1, z_2 \in \Xi_2} \mathcal{G}_1(z_1) + \mathcal{G}_2(z_2) \quad \text{s.t.} \quad \mathbf{A}z_1 + \mathbf{B}z_2 = \mathbf{0}_r,$$

where $\mathbf{A} \in \mathbb{R}^{r \times n_1}$, $\mathbf{B} \in \mathbb{R}^{r \times n_2}$, $\mathcal{G}_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$, $\mathcal{G}_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are closed, proper, and convex functions, and Ξ_1, Ξ_2 are nonempty convex sets. We note that the conditions on \mathcal{G}_1 and \mathcal{G}_2 are pretty mild; in particular, \mathcal{G}_1 and \mathcal{G}_2 are not required to be differentiable and can take on infinite values.

ADMM is an iterative procedure inspired by the well-known method of multipliers [6]. It searches for a minimizer of (B.1) by sequentially minimizing the corresponding augmented Lagrangian with respect to each primal variables z_1 and z_2 , before updating a dual variable $\mathbf{u} \in \mathbb{R}^r$. Formally, the ADMM recursions take the form

$$(B.2) \quad z_1^{(k+1)} = \arg \min_{z_1 \in \Xi_1} \mathcal{G}_1(z_1) + \frac{\rho}{2} \|\mathbf{A}z_1 + \mathbf{B}z_2^{(k)} + \mathbf{u}^{(k)}\|_2^2,$$

$$(B.3) \quad z_2^{(k+1)} = \arg \min_{z_2 \in \Xi_2} \mathcal{G}_2(z_2) + \frac{\rho}{2} \|\mathbf{A}z_1^{(k+1)} + \mathbf{B}z_2 + \mathbf{u}^{(k)}\|_2^2,$$

$$(B.4) \quad \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{A}z_1^{(k+1)} + \mathbf{B}z_2^{(k+1)}$$

for some $\rho > 0$.

ADMM has recently sparked a surge of interest in the signal-processing community for several reasons. First, the conditions on \mathcal{G}_1 and \mathcal{G}_2 in (B.1) (i.e., closed, proper, and convex) are mild and (B.1) therefore encompasses a large number of optimization problems as particular cases. Second, the ADMM recursion (B.2)–(B.4) converges to a solution of (B.1) under very general conditions; see [7, section 3.2]. Third, although ADMM is known to be slow to converge to a solution with high accuracy, it has been shown empirically that ADMM converges to modest accuracy in a few tens of iterations.

Appendix C. Algorithm's details.

C.1. First building block: Computation of (4.5)–(4.7). In this appendix, we complete the exposition done in section 4.1 on the fast evaluation of the gradient of cost function $\mathcal{J}(\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{c})$ given in (4.3), particularized to the model parameters specified in section 5.1. First of all, we expose the particularization of recursions (4.5)–(4.7) to this setting. It is straightforward to see that it results in the following procedure:

(i) Compute sequence $\{\mathbf{x}'_t\}_{t=0}^T$ by the backward recursion:

$$\begin{cases} \mathbf{x}'_T = \mathbf{D}\mathbf{c}', \\ \mathbf{x}'_t = \mathcal{P}(\mathbf{x}'_{t+1}, \mathbf{d}'_{t+1}) + \boldsymbol{\epsilon}'_{t+1}. \end{cases}$$

(ii) Compute sequence $\{\zeta_t\}_{t=0}^T$ by the forward recursion:

$$\begin{cases} \zeta_0 = 2\nabla_{\mathbf{x}_0} \mathcal{H}^*(\mathcal{H}(\mathbf{x}'_0) - \mathbf{y}_0), \\ \zeta_{t+1} = \nabla_{\mathbf{x}_{t+1}} \mathcal{P}^*(\mathbf{x}'_{t+1}, \mathbf{d}'_{t+1})\zeta_t + 2\nabla_{\mathbf{x}_{t+1}} \mathcal{H}^*(\mathcal{H}(\mathbf{x}'_{t+1}) - \mathbf{y}_{t+1}). \end{cases}$$

(iii) Compute the gradients:

$$\begin{cases} \nabla_{\epsilon'} \mathcal{J}(\epsilon', \mathbf{d}', \mathbf{c}') = \zeta_{t-1} + 2\alpha_1 \epsilon'_t, \\ \nabla_{\mathbf{d}'} \mathcal{J}(\epsilon', \mathbf{d}', \mathbf{c}') = \mathbf{E}_t \zeta_{t-1} + 2\alpha_2 \mathbf{W}_t \mathbf{G} \mathbf{G}^* \mathbf{d}'_t, \\ \nabla_{\mathbf{c}'} \mathcal{J}(\epsilon', \mathbf{d}', \mathbf{c}') = \mathbf{D}^* \zeta_T + 2\alpha_3 \mathbf{c}'_t, \end{cases}$$

where $\mathbf{W}_t \in \mathbb{R}^{2n \times 2n}$ and $\mathbf{E}_t \in \mathbb{R}^{2n \times n}$ are, respectively, diagonal and block-diagonal matrices which will be defined in the following. We detail hereafter the elements of the procedure which have not been fully described yet.

We begin by making some comments on the evaluation of the warping function $\mathcal{P}(\mathbf{x}'_t, \mathbf{d}'_t)$ and its Jacobian $\nabla_{\mathbf{x}'_t} \mathcal{P}(\mathbf{x}'_t, \mathbf{d}'_t)$, which constitute the core of the recursion. We propose to use the family of bidimensional cubic cardinal splines $\{\psi_i\}_{i=1}^n$ for the representation (3.3). In practice, we compute an equivalent representation based on the family of bidimensional cubic B-splines functions $\{\phi_i\}_{i=1}^n$. Indeed, this representation presents some computational advantages because of the existence of fast B-splines transforms. The relation between cardinal cubic splines and cubic B-splines functions is given in [52]. This reference also provides details on the fast cubic B-splines transform by recursive filtering. Let matrix $\mathbf{C}^* = [\mathbf{c}_1, \dots, \mathbf{c}_n]^* \in \mathbb{R}^{n \times n}$ denote the direct B-spline transform of a discrete bidimensional signal, i.e., the transform computing from a discrete signal \mathbf{x}_t its representation with spline coefficients $\mathbf{C}^* \mathbf{x}_t$. Rewritten (3.3) with cubic B-spline functions, we get

$$(C.1) \quad \mathcal{P}_s(\mathbf{x}_t, \mathbf{d}_t) = \sum_{i \in \vartheta(\chi(s) + \mathbf{d}_t(s))} \mathbf{c}_i^* \mathbf{x}_t \phi_i(\chi(s) + \mathbf{d}_t(s)),$$

where $\vartheta(\chi(s) + \mathbf{d}_t(s))$ denotes a subset of vector indices corresponding to the neighborhood of the spatial position $\chi(s)$ (which differs from the subset \mathcal{V} previously defined in (3.3)). To simplify notations, we denote by $\mathcal{I} : \mathbb{R}^n \times \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ the function taking as a first argument spline coefficients $\mathbf{C}^* \mathbf{x}_t$ and as a second argument a motion field \mathbf{d}_t , and whose s th component is given by (C.1). Using this notation, (C.1) can be rewritten in the vectorial form

$$\mathcal{P}(\mathbf{x}_t, \mathbf{d}_t) = \mathcal{I}(\mathbf{C}^* \mathbf{x}_t, \mathbf{d}_t).$$

We denote by $\nabla \mathcal{I}(\mathbf{C}^* \mathbf{x}_t, \mathbf{d}_t)$ the Jacobian of function \mathcal{I} at point $(\mathbf{C}^* \mathbf{x}_t, \mathbf{d}_t)$ with respect to its first argument, i.e., spline coefficients. Since function \mathcal{I} is linear with respect to spline coefficients, the Jacobian is only dependent on the value of its second argument, i.e., \mathbf{d}_t . Therefore, we will adopt the notation $\nabla \mathcal{I}(\mathbf{d}_t)$ in what follows.

The complexity of evaluating both spline coefficients $\mathbf{C}^* \mathbf{x}_t$ and the interpolated function \mathcal{I} , scales linearly with the image dimension, i.e., $\mathcal{O}(n)$, thanks to the representation separability and to recursive linear filtering [52]. Multiplication with the Jacobian transpose

$$\nabla_{\mathbf{x}'_t} \mathcal{P}^*(\mathbf{x}'_{t+1}, \mathbf{d}'_{t+1}) = \mathbf{C} \nabla \mathcal{I}^*(\mathbf{d}_{t+1})$$

implies also a linear complexity: first, matrix \mathbf{C} is symmetric¹³ so that it is identical to the direct B-spline transformation \mathbf{C}^* , computed by recursive linear filtering; second, the multiplication of the Jacobian transpose of function \mathcal{I} with vector ζ_t is equal to

$$\nabla \mathcal{I}^*(\mathbf{d}_t)\zeta_t(s) = \sum_{i|s \in \vartheta(\chi(i)+\mathbf{d}_t(i))} \zeta_t(i)\phi_s(\chi(i) + \mathbf{d}_t(i)).$$

Concerning the Jacobian transpose $\nabla_{\mathbf{x}_t}\mathcal{H}^*$, it is easy to see that this matrix is an up-sampling operation, inserting zeros, followed by the same low-pass filtering as in \mathcal{H} .

We continue by detailing matrices appearing in the last step of the procedure. First, we note that matrix \mathbf{D}^* is simply the direct real-valued fast curvelet transform. This transform is, as well as its transpose \mathbf{D} , based on fast Fourier transforms, whose complexity scales in $\mathcal{O}(n \log n)$ [33]. Next, the two diagonals of the two-block matrix \mathbf{E}_t are the two n -dimensional vectors $\partial_{s_j}(\mathcal{I}(\mathbf{C}^*\mathbf{x}_t, \mathbf{d}_t))$ for $j = 1, 2$, where s_j denotes the j th spatial coordinate. We approach these partial derivatives by second-order centered finite differences. Then, the diagonal of matrix \mathbf{W}_t is the vector concatenating twice the weight vector \mathbf{w}_t , i.e., $\mathbf{W}_t(s, s) = \mathbf{W}_t(2s, 2s) = \mathbf{w}_t(s)$ for $s = 1, \dots, n$.

To finalize the description of this procedure, it remains to give some details on matrix \mathbf{G} . Let the elements of vector $\mathbf{G}^*\mathbf{d}_t$ be first-order forward finite difference approximations of the spatial gradients of the two motion components, which have been rearranged beforehand on the pixel grid. This gradient approximation becomes exact assuming that components of vector \mathbf{d}_t are coefficients associated with the decomposition of some continuous motion field in a basis of interpolating and separable scaling functions (see a proof in [31]). Straightforward calculus then shows that elements of vector $\mathbf{G}\mathbf{G}^*\mathbf{d}_t$ are second-order finite difference approximations of the Laplacian of the two motion components, which have been rearranged beforehand on the pixel grid.

C.2. Second building block: ADMM solver for problems (4.17) and (4.19). In this appendix, we present an ADMM implementation of the two minimization problems (4.17) and (4.19) appearing in the procedure described in section 4.3 (which also corresponds to Algorithm 4 introduced later on in section 5). In the following, iterations of the 2-step recursion presented in section 4.3 will be indexed by the exponent (ℓ) , in order to differentiate them from the iterations related to ADMM, which will be indexed by the exponent (k) .

We begin by the analysis of minimization problem (4.17). This problem can be equivalently reexpressed as

$$\begin{aligned} \arg \min_{(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \in \Omega, (\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{c}})} \sum_{t=0}^T \|\mathcal{H}(\mathbf{x}_t) - \mathbf{y}_t\|_2^2 + \sum_{t=1}^T \left(\alpha_1 \|\tilde{\boldsymbol{\epsilon}}_t\|_1 + \gamma \|\tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}\|_1 \right) + \alpha_3 \|\tilde{\mathbf{c}}\|_1 + \gamma \|\tilde{\boldsymbol{\delta}}_{\mathbf{c}}\|_1 \\ \text{s.t.} \quad \begin{cases} \boldsymbol{\epsilon}_t = \tilde{\boldsymbol{\epsilon}}_t & \forall t, \\ \mathbf{H}^*(\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t^{(\ell)}) = \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t} & \forall t, \\ \mathbf{c} = \tilde{\mathbf{c}}, \\ \mathbf{c} - \mathbf{c}^{(\ell)} = \tilde{\boldsymbol{\delta}}_{\mathbf{c}}, \end{cases} \end{aligned}$$

¹³Matrix \mathbf{C} is symmetric in the case of periodic boundary conditions [52].

where

$$\Omega \triangleq \left\{ (\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \left| \begin{array}{l} \mathbf{x}_t = \mathcal{P}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \boldsymbol{\epsilon}_{t+1}, \quad 0 \leq t \leq T-1 \\ \mathbf{x}_T = \mathbf{D}\mathbf{c} \end{array} \right. \right\}.$$

Here, we have added four new variables to the problem, $\tilde{\boldsymbol{\epsilon}} = (\tilde{\boldsymbol{\epsilon}}_1, \dots, \tilde{\boldsymbol{\epsilon}}_T)$, $\tilde{\mathbf{c}}$, $\tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}} = (\tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_1}, \dots, \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_T})$, and $\tilde{\boldsymbol{\delta}}_{\mathbf{c}}$, which are counterbalanced by the inclusion of four new constraints. We use the formalism exposed in Appendix B with $\mathbf{z}_1 = (\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c})$, $\mathbf{z}_2 = (\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{c}}, \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}}, \tilde{\boldsymbol{\delta}}_{\mathbf{c}})$, $\Xi_1 = \Omega$, and $\Xi_2 = \mathbb{R}^{nT} \times \mathbb{R}^q \times \mathbb{R}^{nT} \times \mathbb{R}^q$ and obtain the following ADMM recursions:

$$(C.2) \quad (\mathbf{x}^{(k+1)}, \boldsymbol{\epsilon}^{(k+1)}, \mathbf{c}^{(k+1)}) = \arg \min_{(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) \in \Omega} \mathcal{L}^{(k)}(\mathbf{x}, \boldsymbol{\epsilon}, \mathbf{c}) + \frac{\rho}{2} \sum_{t=1}^T \|\mathbf{H}^*(\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t^{(\ell)}) - \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}^{(k)} + \mathbf{u}_{\boldsymbol{\delta}_{\boldsymbol{\epsilon}_t}}^{(k)}\|_2^2 + \frac{\rho}{2} \|\mathbf{c} - \mathbf{c}^{(\ell)} - \tilde{\boldsymbol{\delta}}_{\mathbf{c}}^{(k)} + \mathbf{u}_{\boldsymbol{\delta}_{\mathbf{c}}}^{(k)}\|_2^2,$$

$$(C.3) \quad \begin{cases} \tilde{\boldsymbol{\epsilon}}_t^{(k+1)} = \arg \min_{\tilde{\boldsymbol{\epsilon}}_t} \|\tilde{\boldsymbol{\epsilon}}_t\|_1 + \frac{\rho_1}{2\alpha_1} \|\boldsymbol{\epsilon}_t^{(k+1)} - \tilde{\boldsymbol{\epsilon}}_t + \mathbf{u}_{\boldsymbol{\epsilon}_t}^{(k)}\|_2^2, \\ \tilde{\mathbf{c}}^{(k+1)} = \arg \min_{\tilde{\mathbf{c}}} \|\tilde{\mathbf{c}}\|_1 + \frac{\rho_3}{2\alpha_3} \|\mathbf{c}^{(k+1)} - \tilde{\mathbf{c}} + \mathbf{u}_{\mathbf{c}}^{(k)}\|_2^2, \\ \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}^{(k+1)} = \arg \min_{\tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}} \|\tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}\|_1 + \frac{\rho}{2\gamma} \|\mathbf{H}^*(\boldsymbol{\epsilon}_t^{(k+1)} - \boldsymbol{\epsilon}_t^{(\ell)}) - \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t} + \mathbf{u}_{\boldsymbol{\delta}_{\boldsymbol{\epsilon}_t}}^{(k)}\|_2^2, \\ \tilde{\boldsymbol{\delta}}_{\mathbf{c}}^{(k+1)} = \arg \min_{\tilde{\boldsymbol{\delta}}_{\mathbf{c}}} \|\tilde{\boldsymbol{\delta}}_{\mathbf{c}}\|_1 + \frac{\rho}{2\gamma} \|\mathbf{c}^{(k+1)} - \mathbf{c}^{(\ell)} - \tilde{\boldsymbol{\delta}}_{\mathbf{c}} + \mathbf{u}_{\boldsymbol{\delta}_{\mathbf{c}}}^{(k)}\|_2^2, \end{cases}$$

$$(C.4) \quad \begin{cases} \mathbf{u}_{\boldsymbol{\epsilon}_t}^{(k+1)} = \mathbf{u}_{\boldsymbol{\epsilon}_t}^{(k)} + \boldsymbol{\epsilon}_t^{(k+1)} - \tilde{\boldsymbol{\epsilon}}_t^{(k+1)}, \\ \mathbf{u}_{\mathbf{c}}^{(k+1)} = \mathbf{u}_{\mathbf{c}}^{(k)} + \mathbf{c}^{(k+1)} - \tilde{\mathbf{c}}^{(k+1)}, \\ \mathbf{u}_{\boldsymbol{\delta}_{\boldsymbol{\epsilon}_t}}^{(k+1)} = \mathbf{u}_{\boldsymbol{\delta}_{\boldsymbol{\epsilon}_t}}^{(k)} + \mathbf{H}^*(\boldsymbol{\epsilon}_t^{(k+1)} - \boldsymbol{\epsilon}_t^{(\ell)}) - \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}^{(k+1)}, \\ \mathbf{u}_{\boldsymbol{\delta}_{\mathbf{c}}}^{(k+1)} = \mathbf{u}_{\boldsymbol{\delta}_{\mathbf{c}}}^{(k)} + \mathbf{c}^{(k+1)} - \mathbf{c}^{(\ell)} - \tilde{\boldsymbol{\delta}}_{\mathbf{c}}^{(k+1)}, \end{cases}$$

where $\mathcal{L}^{(k)}$ is defined in (4.14). Equations (C.2), (C.3), and (C.4) correspond, respectively, to expressions (B.2), (B.3), and (B.4) in Appendix B. We comment on the two first steps of the ADMM algorithm, the last one being trivial. First, as already mentioned, problem (C.2) has the same structural form as the problem addressed in section 4.1. We thus apply the methodology described in section 4.1 to solve this problem via a gradient descent algorithm. The core of this methodology is the computation of the gradient of the cost function with respect to \mathbf{c} and $\boldsymbol{\epsilon}$. The gradient efficient evaluation relies on a backward-forward recursion possessing the structural form of the first building block constituted by (4.5)–(4.7). Some details of the implementation on (4.5)–(4.7) are provided in Appendix C.1 for the particular case of the model parameters given in section 5.1.

We remark that the complexity associated with the evaluation of the gradient scales as $\mathcal{O}(nT + q)$. Second, the optimization problems specified in (C.3) all have simple analytical solutions based on soft-thresholding operators (4.16). We immediately remark that the two first updates in (C.3) are identical to the ADMM steps (4.12) used to treat the convex case in section 4.2. Moreover, the solutions to the last two problems in (C.3) are given by

$$(C.5) \quad \begin{cases} \tilde{\boldsymbol{\delta}}_{\boldsymbol{\epsilon}_t}^{(k+1)}(i) = \text{soft}_{\frac{\gamma}{\rho}} \left(\mathbf{h}_i^*(\boldsymbol{\epsilon}_t^{(k+1)} - \boldsymbol{\epsilon}_t^{(\ell)}) + \mathbf{u}_{\boldsymbol{\delta}_{\boldsymbol{\epsilon}_t}}^{(k)}(i) \right) \\ \tilde{\boldsymbol{\delta}}_{\mathbf{c}}^{(k+1)}(i) = \text{soft}_{\frac{\gamma}{\rho}} \left(\mathbf{c}^{(k+1)}(i) - \mathbf{c}^{(\ell)}(i) + \mathbf{u}_{\boldsymbol{\delta}_{\mathbf{c}}}^{(k)}(i) \right) \end{cases} \quad \forall i,$$

where \mathbf{h}_i is the i th column of \mathbf{H} and “soft” denotes the soft-thresholding operator defined in (4.16).

We continue with the analysis of minimization problem (4.19). We first remark that we can apply the methodology described in section 4.1 to compute the gradient $\nabla_{\mathbf{d}_t} \mathcal{B}(\mathbf{d}^{(\ell)})$ required to build the quadratic approximation (4.20). Once this quadratic approximation has been obtained, the task now is to solve minimization problem (4.19). We can notice that this problem does unfortunately not possess an explicit solution. To circumvent this issue, we use an ADMM strategy, as detailed below. Problem (4.19) is reexpressed as

$$(C.6) \quad \begin{aligned} & \arg \min_{\mathbf{d}, (\tilde{\mathbf{d}}_1, \dots, \tilde{\mathbf{d}}_T)} \mathcal{B}(\mathbf{d}, \mathbf{d}^{(\ell)}) + \alpha_2 \sum_{t=1}^T \mathcal{R}(\tilde{\mathbf{d}}_t) \\ & \text{s.t. } \mathbf{G}^* \mathbf{d}_t = \tilde{\mathbf{d}}_t \quad \forall t. \end{aligned}$$

Here, we have added the new variables $\tilde{\mathbf{d}}_t$'s to the problem which are counterbalanced by the inclusion of new constraints. We use the formalism exposed in Appendix B with $\mathbf{z}_1 = (\mathbf{d}_1, \dots, \mathbf{d}_T)$, $\mathbf{z}_2 = (\tilde{\mathbf{d}}_1, \dots, \tilde{\mathbf{d}}_T)$, $\Xi_1 = \mathbb{R}^{2nT}$, and $\Xi_2 = \mathbb{R}^{hT}$ and obtain the following ADMM recursions:

$$(C.7) \quad \mathbf{d}^{(k+1)} = \arg \min_{\mathbf{d}} \mathcal{B}(\mathbf{d}, \mathbf{d}^{(\ell)}) + \frac{\rho_2}{2} \sum_{t=1}^T \|\mathbf{G}^* \mathbf{d}_t - \tilde{\mathbf{d}}_t^{(k)} + \mathbf{u}_{\mathbf{d}_t}^{(k)}\|_2^2,$$

$$(C.8) \quad \tilde{\mathbf{d}}_t^{(k+1)} = \arg \min_{\tilde{\mathbf{d}}_t} \mathcal{R}(\tilde{\mathbf{d}}_t) + \frac{\rho_2}{2\alpha_2} \|\mathbf{G}^* \mathbf{d}_t^{(k+1)} - \tilde{\mathbf{d}}_t + \mathbf{u}_{\tilde{\mathbf{d}}_t}^{(k)}\|_2^2,$$

$$(C.9) \quad \mathbf{u}_{\mathbf{d}_t}^{(k+1)} = \mathbf{u}_{\mathbf{d}_t}^{(k)} + \mathbf{G}^* \mathbf{d}_t^{(k+1)} - \tilde{\mathbf{d}}_t^{(k+1)}.$$

Equations (C.7), (C.8), and (C.9) correspond, respectively, to expressions (B.2), (B.3), and (B.4) in Appendix B.

We comment now on the resolution of (C.7) and (C.8). First, the unconstrained differentiable problem (C.7) can be easily solved via a gradient descent algorithm. The gradient of the cost function in (C.7) with respect to \mathbf{d}_t can be expressed as

$$(C.10) \quad \nabla_{\mathbf{d}_t} \mathcal{B}(\mathbf{d}^{(\ell)}) + \alpha^{(\ell)} (\mathbf{d}_t - \mathbf{d}_t^{(\ell)}) + \rho_2 \mathbf{G} (\mathbf{G}^* \mathbf{d}_t - \tilde{\mathbf{d}}_t^{(k)} + \mathbf{u}_{\mathbf{d}_t}^{(k)}).$$

As mentioned previously, $\nabla_{\mathbf{d}_t} \mathcal{B}(\mathbf{d}^{(\ell)})$ is simple to evaluate via the recursions described in section 4.1; moreover, the multiplications by \mathbf{G} and \mathbf{G}^* appearing in the last term of (C.10) can be done efficiently for the particular choice of \mathbf{G} considered in this paper (see section C.1 for details on this topic).

Second, the solution of problem (C.8) is closed form (see, e.g., [38, section 6.5.2]). It is given for any $j \in \mathcal{S}_i$ with $1 \leq i \leq n$ by

$$\tilde{\mathbf{d}}_t^{(k+1)}(j) = \begin{cases} 0 & \text{if } \tau_i \leq \alpha_2 \mathbf{w}(i) / \rho_2, \\ \frac{(\tau_i - \alpha_2 \mathbf{w}(i) / \rho_2)}{\tau_i} (\mathbf{g}_j^* \mathbf{d}_t^{(k+1)} + \mathbf{u}_{\tilde{\mathbf{d}}_t}^{(k)}(j)) & \text{otherwise,} \end{cases}$$

where the scalar τ_i is given by

$$\tau_i = \sqrt{\sum_{j \in \mathcal{S}_i} \left(\mathbf{g}_j^* \mathbf{d}_t^{(k+1)} + \mathbf{u}_{\mathbf{d}_t}^{(k)}(j) \right)^2}.$$

Acknowledgments. The authors wish to acknowledge C. Deltel and S. Campion for their technical support in numerical simulations.

REFERENCES

- [1] T. AKGUN, Y. ALTUNBASAK, AND R.M. MERSEREAU, *Super-resolution reconstruction of hyperspectral images*, IEEE Trans. Image Process., 14 (2005), pp. 1860–1875.
- [2] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Ojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.
- [3] H. ATTOUCH, J. BOLTE, AND B. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program., 137 (2013), pp. 91–129.
- [4] S. BAKER AND T. KANADE, *Super-Resolution Optical Flow*, Technical report CMU-RI-TR-99-36, Carnegie Mellon University, Pittsburgh, PA, 1999.
- [5] S. BAKER, D. SCHARSTEIN, J.P. LEWIS, S. ROTH, M. BLACK, AND R. SZELISKI, *A database and evaluation methodology for optical flow*, Int. J. Comput. Vis., 92 (2011), pp. 1–31.
- [6] D.P. BERTSEKAS, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.
- [7] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
- [8] P.J. BURT, *Fast filter transform for image processing*, Comput. Graph. Image Process., 16 (1981), pp. 20–51.
- [9] D.J. BUTLER, J. WULFF, G.B. STANLEY, AND M.J. BLACK, *A naturalistic open source movie for optical flow evaluation*, in European Conference on Computer Vision, Part IV, A. Fitzgibbon et al., eds., Lecture Notes in Comput. Sci. 7577, Springer-Verlag, Berlin, 2012, pp. 611–625.
- [10] E. CANDÈS, L. DEMANET, D. DONOHO, AND L. YING, *Fast discrete curvelet transforms*, SIAM Multiscale Model Simul., 5 (2006), pp. 861–899.
- [11] E.J. CANDÈS AND D.L. DONOHO, *New tight frames of curvelets and optimal representations of objects with piecewise C_2 singularities*, Comm. Pure Appl. Math., 57 (2002), pp. 219–266.
- [12] D. CAPEL AND A. ZISSERMAN, *Super-resolution from multiple views using learnt image models*, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, Vol. 2, IEEE Computer Society, Los Alamitos, CA, 2001, pp. 627–634.
- [13] G.H. COSTA AND J.C.M. BERMUDEZ, *On the design of the LMS algorithm for robustness to outliers in super-resolution video reconstruction*, in 2006 IEEE International Conference on Image Processing, IEEE, Piscataway, NJ, 2006, pp. 1737–1740.
- [14] G.H. COSTA AND J.C.M. BERMUDEZ, *Statistical analysis of the LMS algorithm applied to super-resolution image reconstruction*, IEEE Trans. Signal Process., 55 (2007), pp. 2084–2095.
- [15] G.H. COSTA AND J.C.M. BERMUDEZ, *Registration errors: Are they always bad for super-resolution?*, IEEE Trans. Signal Process., 57 (2009), pp. 3815–3826.
- [16] P. DE SANTIS AND F. GORI, *On an iterative method for super-resolution*, Optica Acta, Internat. J. Opt., 22 (1975), pp. 691–695.
- [17] W. DONG, L. ZHANG, G. SHI, AND X. LI, *Nonlocally centralized sparse representation for image restoration*, IEEE Trans. Image Process., 22 (2013), pp. 1620–1630.

- [18] J. ECKSTEIN AND D.P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.
- [19] M. ELAD AND A. FEUER, *Super-resolution reconstruction of image sequences*, IEEE Trans. Pattern Anal. Mach. Intell., 21 (1999), pp. 817–834.
- [20] M. ELAD AND A. FEUER, *Superresolution restoration of an image sequence: Adaptive filtering approach*, IEEE Trans. Image Process., 8 (1999), pp. 387–395.
- [21] P. ELAD AND A. FEUER, *Super-resolution restoration of continuous image sequence using the LMS algorithm*, in Eighteenth Convention of Electrical and Electronics Engineers in Israel, 1995, IEEE, Piscataway, NJ, 1995, pp. 2.2.5/1–2.2.5/5.
- [22] S. FARSIU, M. ELAD, AND P. MILANFAR, *Video-to-video dynamic super-resolution for grayscale and color sequences*, EURASIP J. Adv. Signal Process., 2006 (2006), 061859.
- [23] S. FARSIU, M.D. ROBINSON, M. ELAD, AND P. MILANFAR, *Fast and robust multiframe super resolution*, IEEE Trans. Image Process., 13 (2004), pp. 1327–1344.
- [24] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Appl. Numer. Harmon. Anal., Birkhäuser, New York, 2013.
- [25] R. FRANSENS, C. STRECHA, AND L. VAN GOOL, *Optical flow based super-resolution: A probabilistic approach*, Comput. Vis. Image Underst., 106 (2007), pp. 106–115.
- [26] R.W. GERCHBERG, *Super-resolution through error energy reduction*, Optica Acta: Internat. J. Opt., 21 (1974), pp. 709–720.
- [27] R. HARDIE, *A fast image super-resolution algorithm using an adaptive Wiener filter*, IEEE Trans. Image Process., 16 (2007), pp. 2953–2964.
- [28] R.C. HARDIE, K.J. BARNARD, J.G. BOGNAR, E.E. ARMSTRONG, AND E.A. WATSON, *High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system*, Opt. Eng., 37 (1998), pp. 247–260.
- [29] L. HE, H. QI, AND R. ZARETZKI, *Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution*, in 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Piscataway, NJ, 2013, pp. 345–352.
- [30] S.H. KELLER, F. LAUZE, AND M. NIELSEN, *Video super-resolution using simultaneous motion and intensity calculations*, IEEE Trans. Image Process., 20 (2011), pp. 1870–1884.
- [31] P. LEMARIÉ-RIEUSSET, *Analyses multirésolutions non orthogonales, commutation entre projecteurs et dérivation et ondelettes vecteurs à divergence nulle*, Rev. Mat. Iberoam., 8 (1992), pp. 221–237.
- [32] C. LIU AND D. SUN, *On Bayesian adaptive video super resolution*, IEEE Trans. Pattern Anal. Mach. Intell., 36 (2014), pp. 346–360.
- [33] S. MALLAT, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic, London, 2008.
- [34] J.M. MENDEL, *Lessons in Estimation Theory for Signal Processing Communications and Control*, Prentice-Hall Signal Process. Ser., Prentice-Hall, Englewood Cliffs, NJ, 1995.
- [35] D. MITZEL, T. POCK, T. SCHOENEMANN, AND D. CREMERS, *Video super resolution using duality based $tv-l_1$ optical flow*, in Pattern Recognition, Lecture Notes in Comput. Sci. 5748, Springer, Berlin, 2009, pp. 432–441.
- [36] K. NASROLLAHI AND T. MOESLUND, *Super-resolution: A comprehensive survey*, Mach. Vis. Appl., 25 (2014), pp. 1423–1468.
- [37] J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization*, Springer Ser. Oper. Res., Springer-Verlag, New York, 1999.
- [38] N. PARIKH AND S. BOYD, *Proximal algorithms*, Found. Trends Optim., 1 (2014), pp. 127–239.
- [39] A.J. PATTI, M.I. SEZAN, AND A.M. TEKALP, *Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time*, IEEE Trans. Image Process., 6 (1997), pp. 1064–1076.
- [40] S. PELEG, D. KEREN, AND L. SCHWEITZER, *Improving image resolution using subpixel motion*, Pattern Recogn. Lett., 5 (1987), pp. 223–226.
- [41] T. PELEG AND M. ELAD, *A statistical prediction model based on sparse representations for single image super-resolution*, IEEE Trans. Image Process., 23 (2014), pp. 2569–2582.
- [42] P. PURKAIT AND B. CHANDA, *Super resolution image reconstruction through Bregman iteration using morphologic regularization*, IEEE Trans. Image Process., 21 (2012), pp. 4029–4039.
- [43] G. PUY AND P. VANDERGHEYNST, *Robust image reconstruction from multiview measurements*, SIAM J. Imaging Sci., 7 (2014), pp. 128–156.

- [44] A. SCHATZBERG AND A.J. DEVANEY, *Super-resolution in diffraction tomography*, Inverse Problems, 8 (1992), pp. 149–164.
- [45] R.R. SCHULTZ AND R.L. STEVENSON, *Extraction of high-resolution frames from video sequences*, IEEE Trans. Image Process., 5 (1996), pp. 996–1011.
- [46] H. TAKEDA AND P. MILANFAR, *Locally adaptive kernel regression for space-time super-resolution*, in Super-Resolution Imaging, Digit. Imaging Comput. Vis., CRC Press, Boca Raton, FL, 2010, pp. 63–69.
- [47] H. TAKEDA, P. MILANFAR, M. PROTTER, AND M. ELAD, *Super-resolution without explicit subpixel motion estimation*, IEEE Trans. Image Process., 18 (2009), pp. 1958–1975.
- [48] B.C. TOM, A.K. KATSAGGELOS, AND N.P. GALATSANOS, *Reconstruction of a high resolution image from registration and restoration of low resolution images*, in Proceedings of the IEEE International Conference on Image Processing, CIP-94, IEEE Computer Society, Los Alamitos, CA, 1994, pp. 553–557.
- [49] R. TSAI AND T. HUANG, *Multiframe image restoration and registration*, Adv. Comput. Vis. Image Process., 1 (1984), pp. 317–339.
- [50] K. TURKOWSKI, *Filters for common resampling tasks*, in Graphics Gems, Academic Press Professional, San Diego, CA, 1990, pp. 147–165.
- [51] M. UNGER, T. POCK, M. WERLBERGER, AND H. BISCHOF, *A convex approach for variational super-resolution*, in Pattern Recognition, M. Goesele, S. Roth, A. Kuijper, B. Schiele, K. Schindler, eds., Lecture Notes Comput. Sci. 6376, Springer, Berlin, 2010, pp. 313–322.
- [52] M. UNSER, A. ALDROUBI, AND M. EDEN, *Fast B-Spline transforms for continuous image representation and interpolation*, IEEE Trans. Pattern Anal. Mach. Intell., 13 (1991), pp. 277–285.
- [53] Z. WANG AND F. QI, *Super-resolution video restoration with model uncertainties*, in Proceedings of the 2002 International Conference on Image Processing, Vol. 2, IEEE, Piscataway, NJ, 2002, pp. 853–856.
- [54] J. WEICKERT AND C. SCHNÖRR, *A theoretical framework for convex regularizers in PDE-based computation of image motion*, Int. J. Comput. Vis., 45 (2004), pp. 245–264.
- [55] L. XU, J. JIA, AND Y. MATSUSHITA, *Motion detail preserving optical flow estimation*, IEEE Trans. Pattern Anal. Mach. Intell., 34 (2012), pp. 1744–1757.
- [56] J. YANG AND D. SCHONFELD, *New results on performance analysis of super-resolution image reconstruction*, in 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, Piscataway, NJ, 2009, pp. 1517–1520.
- [57] J. YANG, J. WRIGHT, T.S. HUANG, AND Y. MA, *Image super-resolution via sparse representation*, IEEE Trans. Image Process., 19 (2010), pp. 2861–2873.
- [58] Q. YUAN, L. ZHANG, AND H. SHEN, *Multiframe super-resolution employing a spatially weighted total variation model*, IEEE Trans. Circuits Syst. Video Technol., 22 (2012), pp. 379–392.
- [59] C. ZACH, T. POCK, AND H. BISCHOF, *A duality based approach for realtime TV- ℓ_1 optical flow*, in Annual Symposium of German Association on Pattern Recognition, Springer, Berlin, 2007, pp. 214–223.
- [60] W. ZHAO AND H. SAWHNEY, *Is super-resolution with optical flow feasible?*, in Computer Vision - ECCV 2002, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, eds., Lecture Notes Comput. Sci. 2350, Springer, Berlin, 2002, pp. 599–613.