Signal Processing 88 (2008) 2890–2901

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Determining the regularization parameters for super-resolution problems ☆

Marcelo V.W. Zibetti [a,*], Fermín S.V. Bazán [b], Joceli Mayer [a]

[a] Department of Electrical Engineering, Federal University of Santa Catarina, 88040-900 Florianópolis, Brazil
[b] Department of Mathematics, Federal University of Santa Catarina, 88040-900 Florianópolis, Brazil

ARTICLE INFO

Article history:
Received 13 October 2007
Received in revised form
3 June 2008
Accepted 14 June 2008
Available online 19 June 2008

Keywords:
Super-resolution
Regularization
Bayesian estimation
JMAP

ABSTRACT

We derive a novel method to determine the parameters for regularized super-resolution problems, addressing both the traditional regularized super-resolution problem with single- and multiple-parameters and the simultaneous super-resolution problem with two parameters. The proposal relies on the joint maximum a posteriori (JMAP) estimation technique. The classical JMAP technique provides solutions at low computational cost, but it may be unstable and presents multiple local minima. We propose to stabilize the JMAP estimation, while achieving a cost function with a unique global solution, by assuming a gamma prior distribution for the hyperparameters. The resulting fidelity is similar to the quality provided by classical methods such as GCV, L-curve and Evidence, which are computationally expensive. Experimental results illustrate the low complexity and stability of the proposed method.

1. Introduction

In many applications it is desirable that the acquisition system provides an image with the best possible resolution while introducing minimum distortions due to imperfections of the image sensor and the optical system. However, the cost of image acquisition systems, like digital cameras, camcorders and scanners, increases with the resolution of the sensor and with the quality of the optical system. An alternative to improve the resolution and the quality of captured images, without increasing the cost of the system, is to employ digital processing techniques to achieve super-resolution (SR).

Research on SR methods dates back to the 90's when the authors in [1,2] employed Fourier domain methods. Since then, many approaches have been proposed, including projections onto convex sets (POCS) [3,4], non-uniform interpolation [5] and iterative back-projection [6,7]. Regularized SR approaches based on maximum a posteriori (MAP) and regularized least squares appeared in [8–11]. In general, regularized approaches are based on the minimization of a cost function composed by a term associated with the residual between the estimated high-resolution (HR) frame and the low-resolution (LR) frames, plus another term, called the prior term, used to regularize the problem. The regularization parameter controls the influence of the prior term in the resulting solution. In many works, the regularization parameter is assumed to be known a priori, but in most cases it is unknown and needs to be determined from the data.

Since regularized SR methods rely on inverse problems theory [12,13], many methods for inverse problems have been used to find the regularization parameter in SR. For instance, the generalized cross validation (GCV), a widely used method for inverse problems [13], was applied to SR in [14]. In [15], the L-curve based method [12], in which the chosen parameter is the one that produces the point of maximum curvature (L-MC) in the L-curve, was proposed

☆ This work was supported by CNPq under Grant nos. 140543/2003-1 and nos. 300487/94 - 0(NV).

* Corresponding author. Tel.: +55 11 30635692.
E-mail addresses: marcelo.zibetti@terra.com.br (M.V.W. Zibetti), fermin@mtm.ufsc.br (F.S.V. Bazán), mayer@eel.ufsc.br (J. Mayer).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.06.010

for SR problems. GCV and L-MC provide high quality solutions, however their computational cost is high. Alternative faster methods are proposed in [16–18]. Other possible methods are those developed for image restoration problems such as [18–20].

One of the difficulties in SR is the existence of motion error between frames. The motion error, caused by imprecise motion estimation or by the occlusion of objects moving in the scene, reduces the effectiveness of SR methods and generates some artifacts in the estimated HR image [21,22]. To overcome this problem, [9,23,24] propose to weight the equations associated to the LR frames independently. However, the choice of proper weighting values to reduce the influence of the frames corrupted by motion errors without completely excluding them from the estimation is a difficult problem. In practice, the weighting values as well as the regularization parameter have to be estimated from the data, which increases the complexity of the problem. The joint determination of the weights and the regularization parameter, simply called multi-parameter problem in this paper, is addressed in [21,22]. In [25], a multi-parameter version of [26], called Evidence, is proposed to solve the problem. The method is stable and provides good quality results. However, it is computationally demanding for general SR problems since, like GCV, it requires computing the trace of the inverse of a matrix. In [25], the method is applied only to block-circulant matrices. Methods based on the gradient, as in [21,22], estimate the HR frame and the parameters at each iteration. These methods have been shown to be stable and are, in general, faster than Evidence but the quality of the estimated frames is inferior.

More recently, simultaneous SR methods have been proposed in [27–30]. These methods estimate all frames of an image sequence in a single process. Two different kinds of prior are employed, one to achieve spatial smoothness and other to achieve higher similarity of the HR frames in the motion trajectory. That is, at least two parameters are necessary, and to the best of our knowledge, a proper method to find the regularization parameters for these techniques has not been proposed yet.

This paper addresses the parameters determination for both the traditional regularized SR problem with single- and multiple-parameters and the simultaneous SR problem with two parameters. These regularized SR algorithms are reviewed in Section 2. In Section 3, the problem of parameters estimation is addressed using the joint maximum a posteriori (JMAP) estimation technique [31,32]. The classical JMAP approach, which assumes uniform density for the hyperparameters is, in general, unstable [32]. This work assumes a gamma probability density for the hyperparameters in order to produce a stable algorithm with a unique global solution. The proposed method is also closely related to the L-curve method in [16] where the chosen parameter produces a point in the L-curve tangent to a line with negative slope. Section 4 presents experiments to illustrate the low computational cost of the proposed method, while producing estimations with the same quality as classical methods such as GCV [14], L-curve [12] and Evidence [25,26]. Section 5 concludes this paper.

2. Review of regularized SR methods and models

In this section, the traditional regularized SR problem and the simultaneous SR problem are reviewed, considering the single- and multi-parameter case for the former and the two-parameter case for the latter.

2.1. Single- and multi-parameter traditional SR

Traditional regularized SR algorithms [33,34] exploit the entire sequence of LR frames to produce a single HR frame. The equation that describes the single-parameter traditional SR method is

$$\hat{\mathbf{f}}_k = \arg\min_{\mathbf{f}_k} \sum_{j=1}^{L} \|\mathbf{g}_j - \mathbf{C}_{j,k}\mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k\mathbf{f}_k\|_2^2. \quad (1)$$

In Eq. (1), $\mathbf{g}_j$ is a vector of size $N \times 1$ that represents the LR frame captured at the time instant j. The elements of the vector correspond to the pixels of the respective frame, lexicographically ordered. The vector $\mathbf{f}_k$, of size $M \times 1$, represents the HR image at instant k, where $N \leqslant M$. The matrix $\mathbf{C}_{j,k} = \mathbf{D}_j\mathbf{M}_{j,k}$ combines the acquisition matrix $\mathbf{D}_j$ and the motion matrix $\mathbf{M}_{j,k}$. The matrix $\mathbf{D}_j$, of size $N \times M$, represents the acquisition process applied to the motion transformed HR image $\mathbf{f}_j$. It models the blurring, caused in the lens and in the image sensor, and the subsampling, which implies the reduction of the number of pixels from the HR frame to the LR frame. The matrix $\mathbf{M}_{j,k}$, of size $M \times M$, represents the motion transformation, or warping. It can be created either from a discretized continuous motion operator [31,35], where a parametric motion is assumed, or from a discrete motion vector field [36,37].

In general, SR is an ill-posed[1] problem [12,13,34,38]. An alternative to achieve a unique and stable solution is to employ a regularization or prior term, represented by $\|\mathbf{R}_k\mathbf{f}_k\|_2^2$ in (1). The matrix $\mathbf{R}_k$, of size $P \times M$, is usually a discrete differential operator, obtained by either employing a finite difference operator (in the horizontal, vertical and diagonals directions) or a Laplacian operator. The regularization parameter, λ_k, which varies according to the temporal position k of the HR frame being estimated, dictates the influence of the prior term in the solution. In this method it is assumed that the data error has the same variance for all frames. Eq. (1) can be represented in condensed form as

$$\hat{\mathbf{f}}_k = \arg\min_{\mathbf{f}_k} \|\boldsymbol{g} - \boldsymbol{C}_k\mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k\mathbf{f}_k\|_2^2 \quad (2)$$

where the data $\boldsymbol{g} = [\mathbf{g}_1^{\mathrm{T}}, \ldots, \mathbf{g}_L^{\mathrm{T}}]^{\mathrm{T}}$ is an $LN \times 1$ vector and $\boldsymbol{C}_k = [\mathbf{C}_{1,k}^{\mathrm{T}} \ldots \mathbf{C}_{L,k}^{\mathrm{T}}]^{\mathrm{T}}$ is an $LN \times M$ matrix.

Multi-parameter traditional SR algorithms also employ the entire sequence of LR frames to produce one HR frame [9,10,24]. However, in these algorithms it is assumed that the noise level on the data is different for each frame specially due to different levels of motion error in each

[1] An ill-posed problem is a mathematical problem that has, at least, one of the following features: it has no solution; it has an infinite number of solutions; or the solution is not stable due to small perturbations in the data [12,13].

frame. Thus, the equations related to the LR frames are weighted individually. The multi-parameter traditional methods are described by

$$\hat{\mathbf{f}}_k = \arg\min_{\mathbf{f}_k} \sum_{j=1}^{L} \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k}\mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k\mathbf{f}_k\|_2^2 \tag{3}$$

where $\alpha_{j,k}$ is the weighting applied to each frame. The parameter $\alpha_{j,k}$ tends to be small with the temporal distance between the frames, $|k-j|$, mainly because of the decreasing of the similarity of the frames.

2.2. Simultaneous SR

The simultaneous algorithms estimate the entire sequence of HR frames in a single process. This approach allows the inclusion of the motion matrix in the prior term, improving the quality of the estimated image sequence. The simultaneous approach was originally proposed in [27], and then improved in [28,29], where the computational cost was reduced by removing the terms with the combined acquisition and motion matrix from the data term. The minimization problem, according to [28], is

$$\begin{aligned}\hat{f}_1, ..., \hat{\mathbf{f}}_L = \arg\min_{\mathbf{f}_1,...,\mathbf{f}_L} & \sum_{k=1}^{L} \|\mathbf{g}_k - \mathbf{D}_k\mathbf{f}_k\|_2^2 + \lambda_R \sum_{k=1}^{L} \|\mathbf{R}_k\mathbf{f}_k\|_2^2 \\ & + \lambda_M \sum_{k=1}^{L-1} \|\mathbf{f}_k - \mathbf{M}_{k,k+1}\mathbf{f}_{k+1}\|_2^2 \end{aligned} \tag{4}$$

where $\|\mathbf{f}_k - \mathbf{M}_{k,k+1}\mathbf{f}_{k+1}\|_2^2$ models the motion difference in the motion trajectory. Eq. (4) uses a first-order finite difference model however second- or arbitrary-order models can also be used [28–30]. Note that the entire HR sequence is estimated simultaneously and that only the acquisition matrix $\mathbf{D}_k$ is utilized in the data term [28–30]. In these methods at least two parameters are necessary. The parameter λ_R controls the spatial smoothness of the images, while λ_M controls the similarity of the estimated images in the motion trajectory. Eq. (4) can be expressed as

$$\begin{aligned}\hat{\boldsymbol{f}} = [\hat{\mathbf{f}}_1^{\mathrm{T}} \dots \hat{\mathbf{f}}_L^{\mathrm{T}}]^{\mathrm{T}} = \arg\min_{\boldsymbol{f}} & \|\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}\|_2^2 + \lambda_R \|\boldsymbol{R}\boldsymbol{f}\|_2^2 \\ & + \lambda_M \|\boldsymbol{M}\boldsymbol{f}\|_2^2 \end{aligned} \tag{5}$$

where $\boldsymbol{g} = [\mathbf{g}_1^{\mathrm{T}} \dots \mathbf{g}_L^{\mathrm{T}}]^{\mathrm{T}}$ is the LR sequence, $\boldsymbol{f} = [\mathbf{f}_1^{\mathrm{T}} \dots \mathbf{f}_L^{\mathrm{T}}]^{\mathrm{T}}$ is the HR sequence, $\boldsymbol{D}$, $\boldsymbol{R}$ are block diagonals defined by $\boldsymbol{D} = \mathrm{diag}(\mathbf{D}_1, \dots, \mathbf{D}_L)$ and $\boldsymbol{R} = \mathrm{diag}(\mathbf{R}_1, \dots, \mathbf{R}_L)$, and

$$\boldsymbol{M} = \begin{bmatrix} \mathbf{I} & -\mathbf{M}_{1,2} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I} & -\mathbf{M}_{L-1,L} \end{bmatrix} \tag{6}$$

for the first-order motion difference, as used in (4), where $\mathbf{I}$ is the identity matrix.

3. Proposed method to estimate the parameters

This section describes the novel approach to estimate the parameters based on the JMAP estimation. A brief review of the classical JMAP for single-parameter problems is provided, followed by the development of the proposed method, optimization schemes, and the extension of the method for problems with multiple-parameters.

3.1. Review of the JMAP estimation

JMAP is a Bayesian estimator that jointly estimates the HR images and the parameters [32,39]. The classical JMAP is presented for single-parameter SR methods according to (2). The time index k of the image, as in $\mathbf{f}_k$, is omitted for clarity. The JMAP estimative is

$$\begin{aligned}\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg\max_{\mathbf{f},\theta,\beta} \rho(\mathbf{f}, \theta, \beta|\boldsymbol{g}) \\ &= \arg\min_{\mathbf{f},\theta,\beta} [-\ln \rho(\boldsymbol{g}|\mathbf{f}, \theta) - \ln \rho(\mathbf{f}|\beta) \\ &\quad - \ln \rho(\theta) - \ln \rho(\beta)] \end{aligned} \tag{7}$$

where $\rho(\mathbf{f}, \theta, \beta|\boldsymbol{g})$ is the posterior density, $\boldsymbol{g}$ is the data vector, or the LR image sequence, $\mathbf{f}$ is a vector representing the HR image, θ is the hyperparameter of the data density $\rho(\boldsymbol{g}|\mathbf{f}, \theta)$ and β is the hyperparameter of the image prior density $\rho(\mathbf{f}\|\beta)$. The functions $\rho(\theta)$ and $\rho(\beta)$ are the prior densities assigned to the hyperparameters, also known as hyperpriors [26,39]. The data density and the image prior density are the same used in the MAP estimations. Let us assume the following Gaussian densities

$$\rho(\boldsymbol{g}|\mathbf{f}, \theta) = \frac{1}{(2\pi\theta)^{LN/2}} \mathrm{e}^{-(\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}\|_2^2/2\theta)} \tag{8}$$

where θ, in this case, is the variance of the data error, and

$$\rho(\mathbf{f}|\beta) = \frac{1}{(2\pi\beta)^{M/2}} \mathrm{e}^{-(\|\mathbf{R}\mathbf{f}\|_2^2/2\beta)}. \tag{9}$$

In this work we assume that θ and β are independent of each other. This assumption implies that the hyperparameter for a class of images of interest is not statistically related to the noise variance of the acquisition system.

In the MAP estimation, the hyperparameters are assumed to have fixed values [39]. Thus, it is not required to estimate them. In this case the regularization parameter is $\lambda = \theta/\beta$. On the other hand, in the JMAP estimation, both the HR images as well as the hyperparameters are random values that need to be estimated from the data. Thus, in the same way that an image prior is needed for the estimation of the HR image, the hyperpriors are needed for the estimation of the hyperparameters.

The approaches in [26,32] assume uniform density for the hyperpriors, so the values are equiprobable, therefore $\rho(\theta) \propto \mathrm{cte}$ and $\rho(\beta) \propto \mathrm{cte}$, for $0 < \theta, \beta < \infty$. The JMAP estimation with these hyperpriors becomes:

$$\begin{aligned}\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg\min_{\mathbf{f},\theta,\beta} & \frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2}{2\theta} + \frac{LN}{2}\ln\theta + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} \\ & + \frac{M}{2}\ln\beta + \mathrm{cte}. \end{aligned} \tag{10}$$

From Eq. (10), it is possible to find the hyperparameters for a fixed $\mathbf{f}$, by differentiating Eq. (10) with respect to the hyperparameters and setting it to zero. This leads to the

following closed form solutions:

$$\hat{\theta} = \frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2}{LN}, \quad \hat{\beta} = \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{M} \tag{11}$$

for the data hyperparameter and for the image hyperparameter, respectively. One can observe that the JMAP estimation of the parameters, with uniform density, is similar to the maximum likelihood (ML) estimation of the parameters [32].

Upon substituting (11) into Eq. (10), results in the following optimization problem:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \ln(\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2) + \frac{M}{LN}\ln(\|\mathbf{R}\mathbf{f}\|_2^2). \tag{12}$$

By determining the gradient of the cost function in (12) the resulting minimizer $\hat{\mathbf{f}}$ is the solution of

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \lambda\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{f} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{g} \tag{13}$$

where λ is

$$\lambda = \frac{\theta}{\beta} = \frac{M}{LN}\frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2}{\|\mathbf{R}\mathbf{f}\|_2^2} \tag{14}$$

This statistical method has great similarity with the deterministic L-Curve method proposed in [16]. An analysis in [16] shows that L-Curve in log–log scale is non-convex. Moreover, the constraining of λ is required to find a proper local minimum. In the Bayesian statistical sense, constrains on λ can be expressed by defining proper hyperparameter priors [26,39]. When employing uniform densities, as done in the classical JMAP, λ is not constrained properly and may result in unstable estimates. More restrictive hyperpriors, on the other hand, may produce a stable estimative and a globally convex problem with a unique minimum.

3.2. Proposed method

The instability of the classical JMAP estimative, according to (12), is reported in [21,26]. An approach to stabilize JMAP by employing a proper hyperprior for general inverse problems is reported in [39]. This work proposes an alternative hyperprior which is able to lead the JMAP to a unique and stable estimative of the parameters. In the JMAP method, the density of the data or the prior density of the images are connected with the density of its respective hyperparameter. For example, the image prior, $\rho(\mathbf{f}|\beta)$, may enforce that the HR image is smooth, constraining the estimative to smooth images. The associated hyperparameter, β, defines "how smooth" is the resulting image. However, when a uniform density is assigned to the hyperparameter, as $\rho(\beta) \propto \mathrm{cte}$, then it is implicitly assumed that an oversmooth image, like a constant intensity value image, when $\beta \to 0$, is as likely to occur as a noisy image, like the one produced by a completely unregulated estimation, when $\beta \to \infty$. Therefore, a more adequate prior density to the hyperparameters is desirable.

A better hyperprior should prevent the hyperparameter to reach very extreme values. The desired prior density for the hyperparameters needs to enforce positive values and provides low probability for very low or very high values. Several candidate densities which present these desirable properties were evaluated for this problem, including gamma, inverse-gamma, log-normal, Maxwell, Rayleigh and Weibull densities. The gamma density, with specific parameters that makes it similar to the chi-squared density, has been shown to have practical and theoretical advantages over the alternatives.

The gamma density provides the constraining necessary to stabilize the problem, making the cost function convex in the entire domain and the solution unique. Besides, the resulting estimation process is quite simple and cheaper, making the analysis less complex. Conceptually, it is shown in [40] that the sum of squared Gaussian random values, as observed in (11), leads to random values with chi-squared density [40], which is a particular case of the gamma density. This point suggests that the gamma is an adequate density to the problem. Thus, the gamma density is proposed for the hyperparameters, with specific parameters that makes it similar to the chi-squared density.

The gamma densities for the hyperparameters are given by

$$\rho(\theta) = \frac{\theta^{a-1}b^{-a}}{\Gamma(a)}\mathrm{e}^{-(\theta/b)}, \quad \rho(\beta) = \frac{\beta^{c-1}d^{-c}}{\Gamma(c)}\mathrm{e}^{-(\beta/d)} \tag{15}$$

where a and c are the scale factors, b and d are the shape factors, and $\Gamma(x)$ is the gamma function [40]. Also, $E\{\theta\} = ab$, $\mathrm{var}\{\theta\} = ab^2$, $E\{\beta\} = cd$ and $\mathrm{var}\{\beta\} = cd^2$.

Substituting the gamma densities in Eq. (7) leads to

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg\min_{\mathbf{f},\theta,\beta} \frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2}{2\theta} + \frac{LN}{2}\ln\theta - (a-1)\ln\theta + \frac{\theta}{b} + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{M}{2}\ln\beta - (c-1)\ln\beta + \frac{\beta}{d} + \mathrm{cte} \tag{16}$$

Note that when $a = LN/2 + 1$ and $c = M/2 + 1$, the gamma density has nearly the same shape as the chi-squared density. These values for a and c provide a necessary condition to achieve a globally convex problem. The b and d will be replaced by expressions involving the expected values of the hyperparameters, namely $b = E\{\theta\}/a = m_\theta/a$ and $d = E\{\beta\}/c = m_\beta/c$. Assigning the mentioned values for a, c, b and d, and applying some algebra, Eq. (16) reduces to

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg\min_{\mathbf{f},\theta,\beta} \frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2}{2\theta} + \frac{\theta(LN+2)}{2m_\theta} + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{\beta(M+2)}{2m_\beta} \tag{17}$$

Differentiating Eq. (17) with respect to the hyperparameters, for fixed $\mathbf{f}$, leads to the following estimative

$$\hat{\theta} = \sqrt{m_\theta}\frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2}{\sqrt{LN+2}}, \quad \hat{\beta} = \sqrt{m_\beta}\frac{\|\mathbf{R}\mathbf{f}\|_2}{\sqrt{M+2}}. \tag{18}$$

Substituting the results of (18) into (17), gives

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \frac{\sqrt{LN+2}}{\sqrt{m_\theta}}\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2 + \frac{\sqrt{M+2}}{\sqrt{m_\beta}}\|\mathbf{R}\mathbf{f}\|_2 \tag{19}$$

which reduces to

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2 + \mu\|\mathbf{R}\mathbf{f}\|_2 \tag{20}$$

where

$$\mu = \sqrt{\frac{m_\theta(M+2)}{m_\beta(LN+2)}} = \sqrt{m_\lambda}\frac{\sqrt{(M+2)}}{\sqrt{(LN+2)}} \tag{21}$$

with m_λ being an average value for λ. Considering the gradient of the cost function in (20) we see that the solution of this optimization problem is found when

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \lambda\mathbf{R}^{\mathrm{T}}\mathbf{R} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{g} \tag{22}$$

where the λ enters as regularization parameter defined by

$$\lambda = \mu\frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2}{\|\mathbf{R}\mathbf{f}\|_2}. \tag{23}$$

In Eq. (23) the value of μ is required. One can determine it by exploiting knowledge about the problem from the average values m_θ and m_β, as shown in Eq. (21). Alternatively, this work proposes to determine μ by performing an analysis of the estimation error. This analysis is provided in Appendix A.1. The experiments presented in Section 4 illustrate the performance of this choice.

The proposed method can also be expressed as the L-curve method, similar to the method proposed in [16], however considering the L-curve in sqrt–sqrt scale. The analysis of the L-curve, also provided in [16], demonstrates that the L-curve in sqrt–sqrt scale is convex.

3.3. Proposed optimization method

This work proposes two optimization methods to find the parameters and the HR images. The first method provides an alternated procedure. Given an initial value λ_0, for $n>0$ the HR image is estimated by

$$\mathbf{f}_n = \arg\min_{\mathbf{f}} \|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}\|_2^2 + \lambda_n\|\mathbf{R}\mathbf{f}\|_2^2 \tag{24}$$

and the minimization process is performed, until convergence, using the iterative linear conjugated gradient (CG) [13,41]. The parameter is updated using

$$\lambda_{n+1} = \mu\frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}_n\|_2}{\|\mathbf{R}\mathbf{f}_n\|_2} \tag{25}$$

and the HR image is re-estimated with the new parameter. This alternated procedure stops when $|\lambda_{n+1} - \lambda_n|/\lambda_n$ is a sufficiently small value. Observe that convergence of the iterative procedure (24) requires the convergence of (25) which is represented by a fixed-point sequence of the form

$$\lambda_{n+1} = \mu\Phi(\lambda_n) \quad \text{with } \Phi(\lambda_n) = \frac{\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}_n\|_2}{\|\mathbf{R}\mathbf{f}_n\|_2} \tag{26}$$

The convergence properties of the fixed-point sequence relies on the fact that $\Phi(\lambda)$ increases with λ. For details, the reader is referred to [42].

The second method provides a direct minimization of the optimization problem (20) using non-linear conjugated gradient (NL-CG), where the parameters and the HR images are updated at each iteration. This procedure is described in Appendix A.2. Convexity of the cost function assures that, at the convergence point, both procedures satisfy (23), indicating that both the optimization problems provide the same solution.

3.4. Extension to multi-parameter problems

The proposed method can be extended to the multi-parameter problems. For the traditional multi-parameter SR in (3), the proposed estimation problem is

$$\hat{\mathbf{f}}_k = \arg\min_{\mathbf{f}_k} \sum_{j=1}^{L} \gamma_{j,k}\|\mathbf{g}_j - \mathbf{C}_{j,k}\mathbf{f}_k\|_2 + \mu_k\|\mathbf{R}_k\mathbf{f}_k\|_2 \tag{27}$$

Observe that only the norm, not the squared norm, is considered in Eq. (27). The associated cost function is also convex. It is obtained by applying similar assumptions used on the single-parameter problem [43]. The equivalent parameters of the proposed method are

$$\alpha_{j,k} = \gamma_{j,k}\frac{\|\mathbf{g}_k - \mathbf{D}_k\mathbf{f}_k\|_2}{\|\mathbf{g}_j - \mathbf{C}_{j,k}\mathbf{f}_k\|_2} \tag{28}$$

for the weighting and

$$\lambda_k = \mu_k\frac{\|\mathbf{g}_k - \mathbf{D}_k\mathbf{f}_k\|_2}{\|\mathbf{R}_k\mathbf{f}_k\|_2} \tag{29}$$

for the regularization parameter. In the Experiments section we illustrate the performance of the method with $\gamma_{j,k}$ and μ_k derived in Appendix A.1.

For the simultaneous SR method, from Eq. (5), the proposed optimization problem to be solved is

$$\hat{\mathbf{f}} = \arg\min_{f} \|\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}\|_2 + \mu_R\|\boldsymbol{R}\boldsymbol{f}\|_2 + \mu_M\|\boldsymbol{M}\boldsymbol{f}\|_2 \tag{30}$$

Similar assumptions about the density of hyperparameters in the JMAP development for the simultaneous problem leads to this proposed method, which also preserves the convexity of the associated cost function. Comparing the solutions of (5) and (30) one can see that the equivalent regularization parameters of the proposed method are

$$\lambda_R = \mu_R\frac{\|\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}\|_2}{\|\boldsymbol{R}\boldsymbol{f}\|_2} \tag{31}$$

for the spatial smoothness term and

$$\lambda_M = \mu_M\frac{\|\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}\|_2}{\|\boldsymbol{M}\boldsymbol{f}\|_2} \tag{32}$$

for the term responsible for the similarity of the HR frames in the motion trajectory. The proposed μ_R and μ_M, used in the experiments of Section 4, are presented in Appendix A.1. The two optimization procedures proposed in Section 3.3 can be employed to solve the problems (27) and (30).

4. Experiments

The following experiment evaluates the performance of the methods in finding the parameters for the SR algorithms discussed in this paper. Given an HR image sequence, with known or previously estimated motion, the simulated acquisition process was performed, employing the average of a squared area of $R \times R$ pixels with subsampling factor of R, where R can be 2 and 3, and an additive white Gaussian noise with variance adjusted to achieve a fixed SNR.[2] Three situations were considered: high acquisition noise, with $\mathrm{SNR_A} = 20\,\mathrm{dB}$; medium noise, with $\mathrm{SNR_A} = 30\,\mathrm{dB}$; low noise, with $\mathrm{SNR_A} = 40\,\mathrm{dB}$. These noise levels are the typical levels found in commercial image sensors[3] [44].

The three SR algorithms reviewed in Section 2 were utilized to recover the HR sequence. For each SR algorithm several available methods were used to determine the parameters. The quality of the HR sequence recovered with the parameters found by a particular method is measured in terms of SNR [19]. Computational effort of each method was evaluated by considering the time it takes for convergence, where convergence is assumed to be reached when the improvement in quality is less than 10^{-2} dB. This procedure was repeated using 20 random realizations of the noise, for each noise level. The entire experiment was repeated for each image sequence of a total of six different image sequences. In some of the sequences, the motion was artificially generated without considering occlusions in the scene, whereas in other sequences, which are from real video sequences, the motion was estimated using the optical flow method [45]. In this case, linear interpolated versions of the LR images were employed. The estimated motion vectors are not completely reliable in this case, therefore, occlusions and motion errors occur in several places in the sequence. In this evaluation, the procedure of detection and removal of the occlusion regions was not considered in order to evaluate the performance of the methods in finding the proper parameters to reduce the distortions caused by these errors.

At the end of this section, some visual experiments applying SR in real video sequences are performed, without simulated acquisition. In this case, a procedure of detection and removal of the occlusion regions was considered [46] to achieve the best visual quality.

The methods used to find the parameters are mentioned below. All of them are used in the single-parameter traditional SR and some of them in conjunction with multi-parameters traditional SR. Concerning the parameter method for simultaneous SR, only the classical JMAP and the proposed method are utilized. The implemented methods are:

GCV: Generalized cross validation, as described in [14].

L-MC: An L-curve method, where the parameter provides the point of maximum curvature as proposed in [15].

K-HE: A deterministic method proposed in [22].

EVID: The statistical method Evidence, proposed in [25]. To apply this method to non block-circulant matrices, the trace of the inverse matrix is statistically estimated using the same procedure as in [14].

JMAP: The classical JMAP approach [26,32] as Eq. (12), using CG to find the HR images with (11) to update the parameters.

PROP-1: Proposed method using the alternated procedure with CG to find the HR images and with (25) to update the parameters.

PROP-2: Proposed method with direct minimization using NL-CG as presented in Appendix A.2.

All these methods are iterative. The CG method is used in GCV, L-MC, EVID, JMAP and in the proposed method to find the HR images. K-HE is limited to the gradient method. The same initial conditions are considered: the initial HR image is a null image, and the initial parameters is randomly chosen from 10^{-6} to 10^{6}. Beside these methods, the results obtained by the following predetermined parameters were also compared:

KNOWN: Employs the MAP estimative where the parameters are known *a priori*. Since the noise and the original HR images are known in the experiments, the hyperparameters are computed without difficulties. This method is used as reference only, since it cannot be used in practice.

INTERP: Estimative with the parameters obtained from applying the maximum likelihood in Eq. (11), employing an interpolated version of the LR images. This approach assumes that the interpolated image is a good substitute for the original HR image in order to find the hyperparameters.

4.1. Experiments with single-parameter traditional SR algorithm

Average quality, measured by SNR, standard deviation (STD) and relative computational cost of estimated images and its parameters using each method are shown in Table 1. Relative computational cost, denoted by CT, is expressed in terms of the ratio of the cost (in time) of each method to that required by KNOWN. Visual results are presented in Figs. 1 and 2.

Table 1
Average of the SNR, in dB, the standard deviation (STD) and the relative computational time (CT) for the single-parameter traditional SR algorithm

Method	$R=2$, $\mathrm{SNR_A}=20\,\mathrm{dB}$			$R=2$, $\mathrm{SNR_A}=40\,\mathrm{dB}$			$R=3$, $\mathrm{SNR_A}=30\,\mathrm{dB}$		
	SNR	STD	CT	SNR	STD	CT	SNR	STD	CT
GCV	20.8	1.8	63.9	22.7	2.6	77.6	18.0	1.1	141.1
L-MC	20.3	1.6	58.0	23.3	1.4	84.9	19.1	1.7	243.2
K-HE	18.7	0.9	21.3	20.9	1.0	35.5	16.8	0.9	53.2
EVID	20.2	1.1	52.1	23.1	1.5	51.5	18.9	1.2	231.7
JMAP	10.9	10.1	7.2	7.9	12.2	11.4	13.3	4.9	47.1
KNOWN	19.8	0.9	1.0	23.4	0.9	1.0	18.3	1.0	1.0
INTERP	18.1	1.0	1.2	20.2	1.1	1.6	17.4	1.6	5.1
PROP-1	20.8	1.0	6.9	23.3	1.3	10.1	20.0	1.8	31.2
PROP-2	20.8	1.0	1.1	23.3	1.3	1.2	20.0	1.8	2.7

[2] The acquisition SNR is defined as $\mathrm{SNR_A} = 10\log_{10}(\sigma^2_{Df}/\sigma^2_{\eta})$, where σ^2_{Df} is an LR noise-free sequence variance and σ^2_{η} is the noise variance.

[3] Typical acquisition SNR may vary from 10 to 40 dB, depending on the exposure [44].

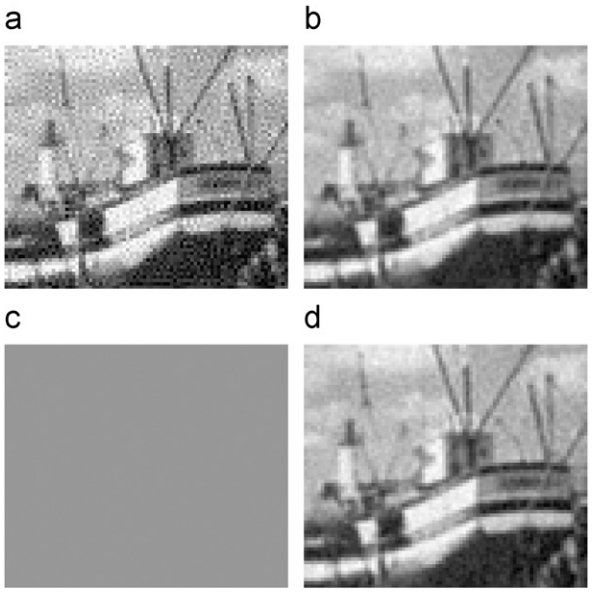

Fig. 1. Visual results of an image of the sequence Boat, (a) JMAP (SNR = 18.1 dB) and (b) PROP-1 (SNR = 23.4 dB) from one realization and and (c) JMAP (SNR = 7.4 dB) and (d) PROP-1 (SNR = 23.8 dB) from a different realization, with $R = 2$ and $\mathrm{SNR_A} = 20$ dB.

Table 1 shows average quality results (SNR) and corresponding STD for six different sequences and 20 realizations. One can observe that the quality of the proposed method is as good as the quality obtained by the best classical methods, such as GCV, L-MC and EVID. The high STD illustrates that the JMAP method is unstable, as reported in [21,26]; during the experiments it diverged many times.

Fig. 1 illustrates the instability problem of the classical JMAP approach. Note that the JMAP result depicted in Fig. 1(a) seems reasonable, but in Fig. 1(c), that shows the results for another realization, it is seen that JMAP diverges producing a constant intensity image, which is unacceptable for a SR method. The proposed method, on the other hand, is very stable. The visual results present the same level of smoothness in Figs. 1(b) and (d).

Fig. 2 presents the results for a sequence with real video, which presents motion errors. In this case, the regularization parameter needs to provide enough smoothness in the image to avoid the amplification of noise and the motion errors. The distortions caused by the motion errors occur, mainly, around the tree in this scene. Note that all parameter determination methods enforce smoothness around the tree which results in blurring of this region. One way to reduce the problem caused by the motion errors in the traditional SR method is by the proper weighting of the LR frames. This is done in the multi-parameter traditional SR method.

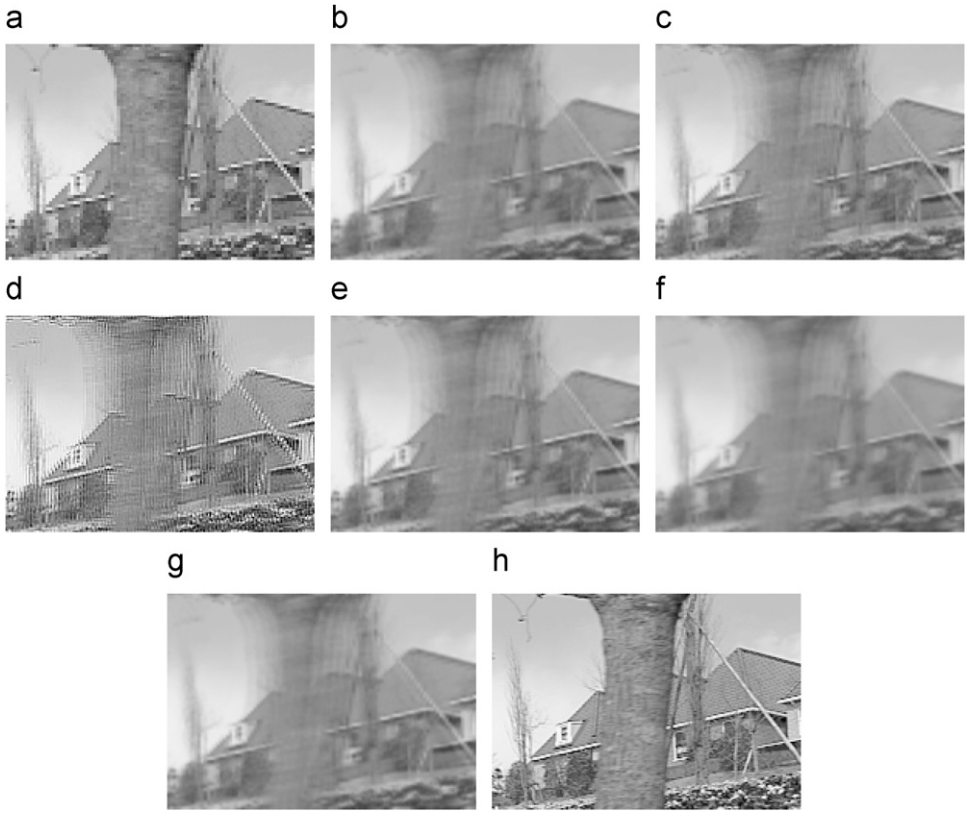

Fig. 2. Visual result of an image of the sequence Flower Garden, with $R = 2$ and $\mathrm{SNR_A} = 40$ dB. (a) Captured image, (b) GCV (SNR = 15.8 dB), (c) L-MC (SNR = 16.2 dB), (d) K-HE (SNR = 14.6 dB), (e) EVID (SNR = 16.1 dB), (f) KNOWN (SNR = 15.5 dB), (g) PROP-1 (SNR = 16.0 dB) and (h) original image.

Table 1 also illustrates the computational cost of each method. The results illustrate that the proposed method PROP-1 provides lower computational cost than the existing methods, with cost similar to the one provided by the classical JMAP approach. Moreover, the proposed method PROP-2, that utilizes direct minimization through NL-CG, provides even lower computational cost. The cost of the proposed method PROP-2 is similar to the cost of a CG minimization with fixed parameters.

4.2. Experiments with multi-parameter traditional SR algorithm

The average quality of the estimated images, its STD and the relative computational time together with the parameters found by the respective method are shown in Table 2. Some visual results are shown in the Fig. 3.

Table 2 shows that the quality obtained by the proposed method is satisfactory and similar to the results of KNOWN. Also, the computational cost results illustrate the low computational cost provided by the proposed method. The performance of the proposed method was superior than the K-HE, recently developed for the multi-parameter traditional SR.

Fig. 3 illustrates the performance of the methods in controlling the weighting in order to avoid the distortions caused by the large motion errors. One can see that the results of the proposed method were very similar to the results of KNOWN. Also, in this example, one can see that the distortions caused by the occlusions were not completely removed, but they were significantly attenuated. The complete removal of these distortions requires the use of a robust SR method [47], or an occlusion and motion error detection and removal procedure [46].

Table 2
Average of the SNR, in dB, the standard deviation (STD) and the relative computational time (CT) for multi-parameter traditional SR algorithm

Method	$R=2$, $SNR_A=20$ dB			$R=2$, $SNR_A=40$ dB			$R=3$, $SNR_A=30$ dB		
	SNR	STD	CT	SNR	STD	CT	SNR	STD	CT
K-HE	16.5	1.9	20.6	17.8	2.2	21.6	14.0	2.4	46.0
EVID	20.8	1.6	90.2	23.7	2.5	190.0	19.3	1.1	466.1
JMAP	15.3	6.1	66.5	18.4	6.2	58.6	11.6	3.6	84.3
KNOWN	22.0	0.6	1.0	26.6	0.5	1.0	20.9	1.1	1.0
INTERP	14.7	0.6	3.1	14.5	0.6	3.0	14.5	1.3	12.1
PROP-1	22.7	0.9	15.1	26.1	1.6	16.7	21.2	1.9	31.2
PROP-2	22.7	0.9	1.8	26.1	1.6	4.3	21.2	1.9	2.7

4.3. Experiments with two-parameter simultaneous SR algorithm

The average quality of the estimated images with the parameters found by the respective method, its STD, and the relative computational time are shown in Table 3. One can observe in Table 3 that the quality obtained by the proposed method was superior to the obtained by KNOWN. As far as the authors know, besides the JMAP method, there is no other method to determine the parameters for the simultaneous SR methods in order to compare with the proposed method. Note that the JMAP method was more stable with the simultaneous SR method than with the traditional SR methods, showing lower STD.

The results from Table 3 also illustrate the computational cost of the methods. The classical JMAP approach was very fast with the simultaneous SR methods and faster than PROP-1. However, the proposed method PROP-2 was even faster. The computational cost of PROP-2 is comparable with the cost taken by the CG with fixed parameters, such as KNOWN or INTERP.

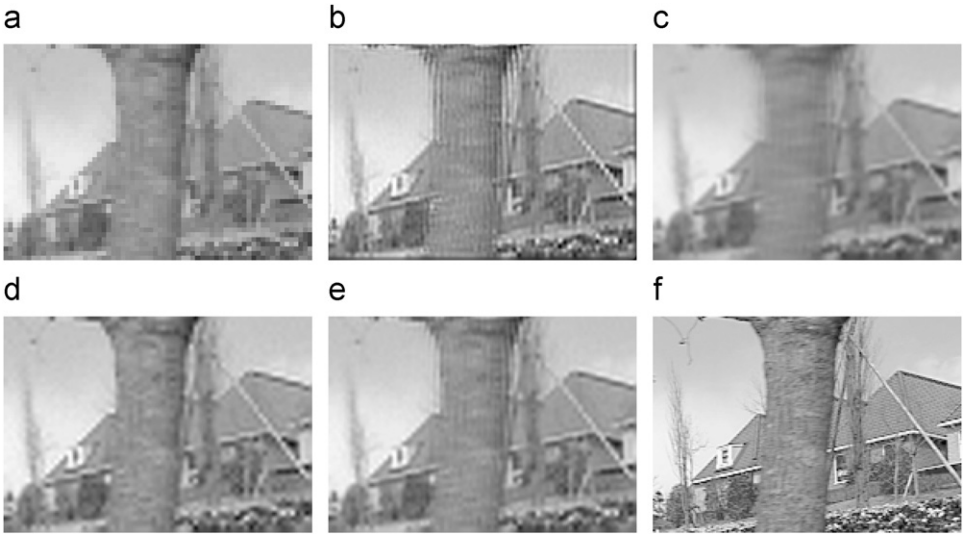

Fig. 3. Visual results from an image of the sequence Flower Garden, with $R=3$ and $SNR_A=30$ dB. (a) Captured image, (b) K-HE (SNR = 14.7 dB), (c) EVID (SNR = 16.2 dB), (d) KNOWN (SNR = 17.1 dB), (e) PROP-1 (SNR = 17.2 dB) and (f) original image.

Table 3
Average of the SNR, in dB, the standard deviation (STD) and the relative computational time (CT) for the two-parameter simultaneous SR algorithm

Method	$R=2, \mathrm{SNR_A}=20$ dB			$R=2, \mathrm{SNR_A}=40$ dB			$R=3, \mathrm{SNR_A}=30$ dB		
	SNR	STD	CT	SNR	STD	CT	SNR	STD	CT
JMAP	22.6	1.1	5.8	24.0	1.2	3.9	20.0	2.0	8.1
KNOWN	22.1	0.4	1.0	26.4	0.5	1.0	20.8	1.2	1.0
INTERP	18.3	0.4	0.7	19.2	0.6	0.6	16.6	1.4	1.3
PROP-1	23.1	0.4	13.0	26.8	0.6	12.0	21.2	2.0	25.7
PROP-2	23.1	0.4	1.3	26.8	0.6	2.0	21.2	2.0	3.1

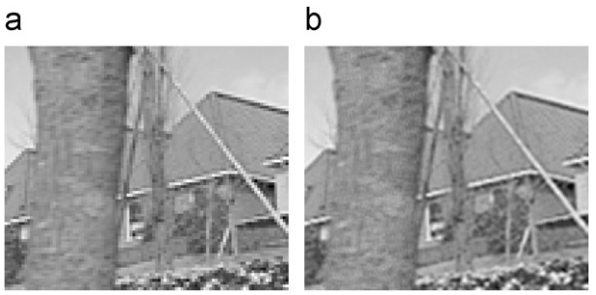

Fig. 4. Visual results comparing the original frame of the sequence Flower Garden with the same frame, where a resolution improvement factor of $R=3$ was applied. (a) Image at original resolution and (b) SR over the original.

4.4. Example with practical SR algorithm

Fig. 4 shows a result without artificial degradation and recovering, using $R=3$. For this visual experiment, the original sequence is assumed to be the captured sequence and used as test problem for the simultaneous SR method with occlusion and motion error detection and removal [46] procedures. The enhancement of the resolution of the SR method over the original image can be clearly noticed. The regularization parameter was determined by the proposed method.

5. Conclusions

In this paper, a technique to determine the parameters for super-resolution methods is proposed. The proposed technique can be applied to the traditional regularized super-resolution problem with single parameter and with multiple parameters, and also to the simultaneous super-resolution problem with two parameters. The problem of parameters estimation has been addressed with the Bayesian theory, using joint maximum a posteriori (JMAP) estimation. A gamma density is proposed for the hyper-parameters in order to provide a globally convex cost function, resulting in a unique solution. The proposed method is also similar to the L-curve methods in [16], however with sqrt–sqrt scale, which is a convex function. The proposed method provides very low computational cost and produces estimated images with the same quality as the ones provided by classical methods. We provide a set of experiments to illustrate the superior efficiency and stability of the proposed technique when compared with other competing methods.

Appendix A

A.1. Alternative choice of parameter μ

This section presents an alternative choice of μ based on the analysis of the estimation error. The advantage of this choice is that it does not depend on the knowledge of m_θ and m_β, as described in Eq. (21).

According to inverse problems theory [12,13], the estimation error, $\mathbf{e}_\lambda$, can be split into three components

$$\mathbf{e}_\lambda = \mathbf{f} - \mathbf{f}_\lambda = \mathbf{e}_{\mathrm{cte}} + \mathbf{e}_\eta(\lambda) + \mathbf{e}_\mathrm{s}(\lambda) \tag{A.1}$$

Here $\mathbf{e}_{\mathrm{cte}}$ is an error vector that does not depend on λ, $\mathbf{e}_\eta(\lambda)$ is the error caused by the amplification of the noise, which is significant when λ is very small (smaller than the optimal value) and $\mathbf{e}_\mathrm{s}(\lambda)$ is the error caused by excessive regularization (oversmoothing), which becomes dominant when λ is very large (larger than the optimal value). The estimation error satisfies $\|\mathbf{e}_\lambda\|_2 \leqslant \|\mathbf{e}_{\mathrm{cte}}\|_2 + \|\mathbf{e}_\eta(\lambda)\|_2 + \|\mathbf{e}_\mathrm{s}(\lambda)\|_2$, and the assumption used through the analysis is that the minimizer of $\|\mathbf{e}_\lambda\|_2$ is rather close to the minimizer of $\|\mathbf{e}_\eta(\lambda)\|_2 + \|\mathbf{e}_\mathrm{s}(\lambda)\|_2$, as illustrated further in the Fig. A1(c). The choice of μ will then result as a consequence of relating $\|\mathbf{e}_\eta(\lambda)\|_2$ to $\|\mathbf{R}\mathbf{f}_\lambda\|_2$ and $\|\mathbf{e}_\mathrm{s}(\lambda)\|_2$ to $\|\boldsymbol{g} - \boldsymbol{C}\mathbf{f}_\lambda\|_2$. The key tool for the analysis is the generalized singular value decomposition (GSVD).

Given the matrices $\boldsymbol{C}$ of size $LN \times M$, and $\mathbf{R}$ of size $P \times M$ with $LN \geqslant M \geqslant P$, the GSVD of the pair $(\boldsymbol{C}, \mathbf{R})$ reads

$$\boldsymbol{C} = \sum_{i=1}^{M} \mathbf{u}_i \sigma_i \mathbf{x}_i^{\mathrm{T}}, \quad \mathbf{R} = \sum_{i=1}^{P} \mathbf{v}_i v_i \mathbf{x}_i^{\mathrm{T}} \tag{A.2}$$

where the $\mathbf{u}_i$ and $\mathbf{v}_i$ are column vectors of orthonormal matrices, and $\mathbf{x}_i^{\mathrm{T}}$ are the rows of a non singular matrix $\mathbf{X}$ with inverse $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M]$. The scalars v_i are ordered so that $1 > v_1 \geqslant v_2 \geqslant \cdots \geqslant v_P > 0$ and the scalars σ_i are ordered so that $0 < \sigma_1 \leqslant \cdots \leqslant \sigma_P < \sigma_{P+1} = \cdots = \sigma_M = 1$. The generalized singular values are $\gamma_i = \sigma_i / v_i$, for $i = 1, \ldots, P$. Further details about the GSVD can be found in [12].

Based on the GSVD, the vectors that compose the error can be expressed as

$$\mathbf{e}_\mathrm{s}(\lambda) = \sum_{i=1}^{P} \left(\frac{\lambda}{\gamma_i^2 + \lambda} \right) (\mathbf{x}_i^{\mathrm{T}} \mathbf{f}) \mathbf{y}_i \tag{A.3}$$

and

$$\mathbf{e}_\eta(\lambda) = -\sum_{i=1}^{P} \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \right) \left(\frac{\eta_i}{\sigma_i} \right) \mathbf{y}_i \tag{A.4}$$

where $\boldsymbol{\eta}$ stands for noise on the data and $\eta_i = \mathbf{u}_i^{\mathrm{T}} \boldsymbol{\eta}$. Additionally, if $\mathbf{d}(\lambda) = \boldsymbol{g} - \boldsymbol{C}\mathbf{f}_\lambda$ and $\mathbf{w}(\lambda) = \mathbf{R}\mathbf{f}_\lambda$, using the GSVD it follows that

$$\mathbf{d}(\lambda) = \sum_{i=1}^{P} \left(\frac{\lambda}{\gamma_i^2 + \lambda} \right) (\sigma_i (\mathbf{x}_i^{\mathrm{T}} \mathbf{f}) + \eta_i) \mathbf{u}_i + \eta_\perp \tag{A.5}$$

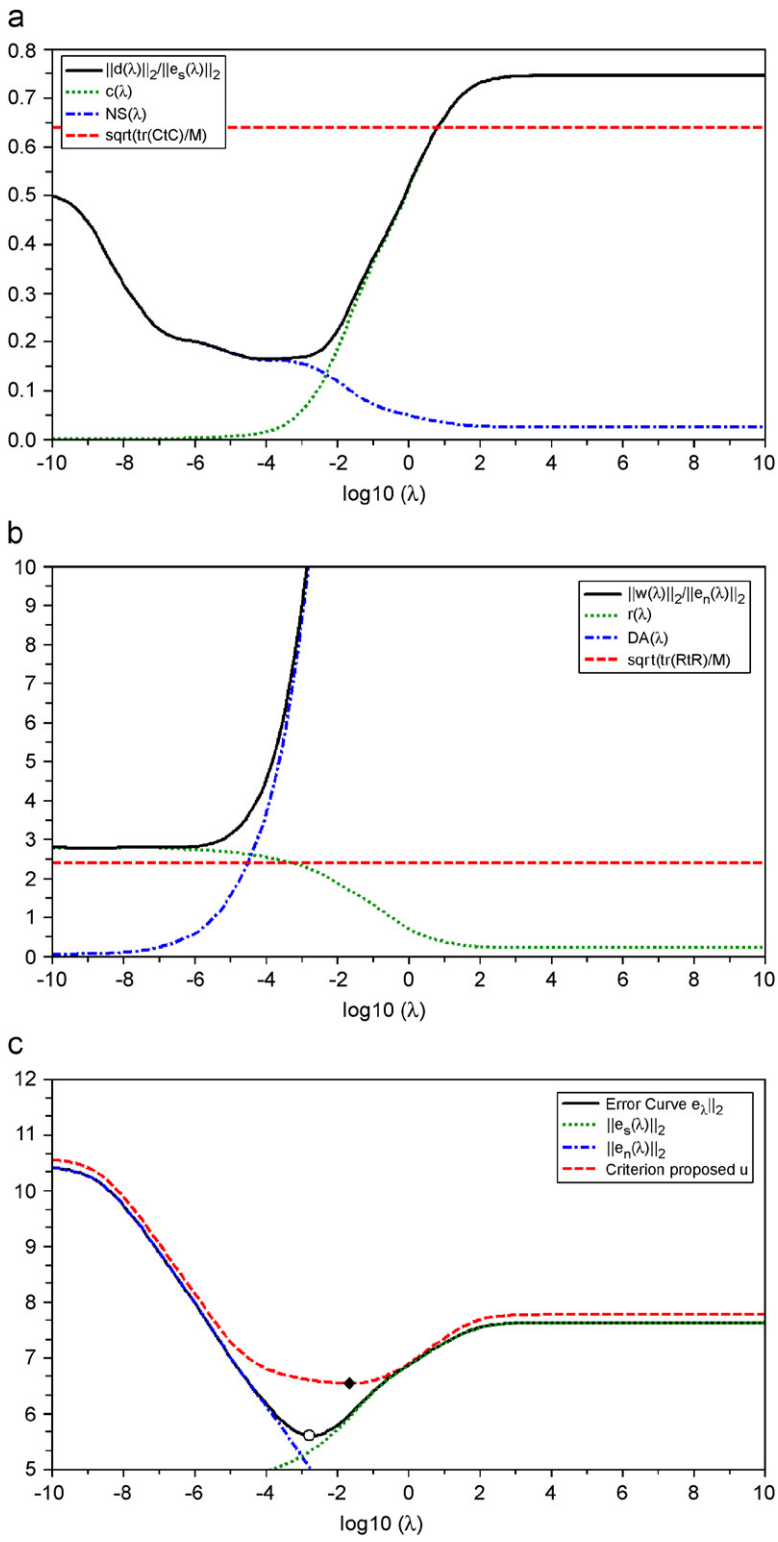

Fig. A1. Graphic illustration of analysis concerning the choice of μ. (a) Behavior of $\|\mathbf{d}(\lambda)\|_2/\|\mathbf{e}_s(\lambda)\|_2$, (b) behavior of $\|\mathbf{w}(\lambda)\|_2/\|\mathbf{e}_\eta(\lambda)\|_2$, and (c) behavior of $\|\mathbf{e}_\lambda\|_2$ and $\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}_\lambda\|_2+\mu\|\mathbf{R}\mathbf{f}_\lambda\|_2$ with μ according to (A.16), also $\|\mathbf{e}_s(\lambda)\|_2$ and $\|\mathbf{e}_\eta(\lambda)\|_2$.

and

$$\mathbf{w}(\lambda)=\sum_{i=1}^{P}\left(\frac{\gamma_i^2}{\gamma_i^2+\lambda}\right)\left(v_i(\mathbf{x}_i^{\mathrm{T}}\mathbf{f})+\frac{v_i\eta_i}{\sigma_i}\right)\mathbf{v}_i \tag{A.6}$$

Comparison of (A.3) with (A.5), and (A.4) with (A.6) leads to

$$\mathbf{d}(\lambda)=\boldsymbol{C}\mathbf{e}_{\mathrm{s}}(\lambda)+\eta_{\mathrm{FILT}}(\lambda)+\eta_{\perp} \tag{A.7}$$

where $\eta_{\mathrm{FILT}}(\lambda)=\sum_{i=1}^{P}(\lambda/(\gamma_i^2+\lambda))(\eta_i)\mathbf{u}_i$ and

$$\mathbf{w}(\lambda)=-\mathbf{R}\mathbf{e}_{\eta}(\lambda)+\mathbf{R}\mathbf{f}_{\mathrm{FILT}}(\lambda) \tag{A.8}$$

where $\mathbf{f}_{\mathrm{FILT}}(\lambda)=\sum_{i=1}^{P}(\gamma_i^2/(\gamma_i^2+\lambda))(\mathbf{x}_i^{\mathrm{T}}\mathbf{f})\mathbf{y}_i$.

In order to relate $\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2$ to $\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}_{\lambda}\|_2$ the following approximation is used

$$\|\mathbf{d}(\lambda)\|_2\approx\|\boldsymbol{C}\mathbf{e}_{\mathrm{s}}(\lambda)\|_2+\|\eta_{\mathrm{FILT}}(\lambda)\|_2+\|\eta_{\perp}\|_2 \tag{A.9}$$

From this it follows that

$$\|\mathbf{d}(\lambda)\|_2/\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2\approx c(\lambda)+NS(\lambda) \tag{A.10}$$

where $NS(\lambda)=(\|\eta_{\mathrm{FILT}}(\lambda)\|_2+\|\eta_{\perp}\|_2)/\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2$ and $c(\lambda)=\|\boldsymbol{C}\mathbf{e}_{\mathrm{s}}(\lambda)\|_2/\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2$. Observe that the Rayleigh quotient of $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$ guarantees that $c(\lambda)^2$ is necessarily between the smallest and the largest eigenvalue of $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$. The approximation considered in (A.10) is not good in general, but it is quite acceptable in SR problems, as illustrated in Fig. A1(a). Further, since the excessive regularization error becomes dominant for large λ, $NS(\lambda)$ gets small and (A.10) reduces approximately to

$$\|\mathbf{d}(\lambda)\|_2/\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2\approx c(\lambda) \tag{A.11}$$

Concerning this approximation, exhaustive numerical simulations showed that $c(\lambda)^2$ behaves nearly as a constant which can be roughly approximated by the mean of the eigenvalues of $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$, obtained through $\sqrt{\mathrm{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})/M}$, where $\mathrm{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})$ is the trace of the matrix $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$, see Fig. A1(a). Thus for large λ the approximation (A.11) can be rewritten as

$$\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2\approx\sqrt{M}\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}_{\lambda}\|_2/\sqrt{\mathrm{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})}. \tag{A.12}$$

Proceeding as before one can see that for small λ

$$\|\mathbf{w}(\lambda)\|_2/\|\mathbf{e}_{\eta}(\lambda)\|_2\approx r(\lambda)+DA(\lambda) \tag{A.13}$$

where $r(\lambda)=\|\mathbf{R}\mathbf{e}_{\eta}(\lambda)\|_2/\|\mathbf{e}_{\eta}(\lambda)\|_2$, with $r(\lambda)^2$ ranging between the smallest and the largest eigenvalue of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$, and $DA(\lambda)=\|\mathbf{R}\mathbf{f}_{\mathrm{FILT}}(\lambda)\|_2/\|\mathbf{e}_{\eta}(\lambda)\|_2$. Since the amplified noise error $\|\mathbf{e}_{\eta}(\lambda)\|_2$ is predominant when λ is very small, thereby implying that $DA(\lambda)$ gets small, the above approximation can also be simplified to yield $\|\mathbf{w}(\lambda)\|_2/\|\mathbf{e}_{\eta}(\lambda)\|_2\approx r(\lambda)$. Numerical simulations showed that for small λ, $r(\lambda)$ behaves approximately as a constant which can be roughly approximated by the mean of the eigenvalues of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$, see Fig. A1(b). This results in:

$$\|\mathbf{e}_{\eta}(\lambda)\|_2\approx\sqrt{M}\|\mathbf{R}\mathbf{f}_{\lambda}\|_2/\sqrt{\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R})} \tag{A.14}$$

Finally, since $\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2$ dominates $\|\mathbf{e}_{\eta}(\lambda)\|_2$ for large λ, and $\|\mathbf{e}_{\eta}(\lambda)\|_2$ dominates $\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2$ for small λ, taking into account (A.12) and (A.14) it is reasonable to expect that for all λ

$$\|\mathbf{e}_{\mathrm{s}}(\lambda)\|_2+\|\mathbf{e}_{\eta}(\lambda)\|_2\approx\mathrm{cte}(\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}_{\lambda}\|_2+\mu\|\mathbf{R}\mathbf{f}_{\lambda}\|_2) \tag{A.15}$$

with

$$\mu=\sqrt{\mathrm{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})}/\sqrt{\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R})} \tag{A.16}$$

The effect of μ is illustrated in Fig. A1(c). Observe that, with the proposed μ, the curve $(\lambda,\|\boldsymbol{g}-\boldsymbol{C}\mathbf{f}_{\lambda}\|_2+\mu\|\mathbf{R}\mathbf{f}_{\lambda}\|_2)$ reaches its minimum (black diamond mark) at a point very close to the minimum (small circle) of the error curve $(\lambda,\|\mathbf{e}_{\lambda}\|_2)$, which justifies the choice of μ according to (A.16).

A.1.1. Extension for multi-parameter problems

Following the ideas of the single-parameter problem, the parameters $\gamma_{j,k}$ and μ_k used in (27) can be chosen as

$$\gamma_{j,k}=1/(1+|j-k|) \tag{A.17}$$

to reduce the influence of the LR frames far apart in time, and

$$\mu_k=\sqrt{\mathrm{tr}(\boldsymbol{D}_k^{\mathrm{T}}\boldsymbol{D}_k)}/\sqrt{\mathrm{tr}(\mathbf{R}_k^{\mathrm{T}}\mathbf{R}_k)} \tag{A.18}$$

Parameters μ_R and μ_M used in (30) can be chosen as

$$\mu_R=\sqrt{\mathrm{tr}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})}/\sqrt{2\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R})} \tag{A.19}$$

and

$$\mu_M=\sqrt{\mathrm{tr}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})}/\sqrt{2\mathrm{tr}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})} \tag{A.20}$$

These suggestions are used in the experiments of Section 4.

A.2. Non-linear conjugate gradient

The direct minimization of Eq. (20) can be performed efficiently using NL-CG [13,41], presented in Table A1

The method stops when a specified number of iterations n is reached or when e_{n+1} becomes lower than

Table A1
Non-linear conjugate gradient

$n:=0$;	
$\mathbf{f}_0:=$ initial HR image guess	
$\lambda_0:=$ initial parameter guess	
$\mathbf{r}_0:=\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{C}\mathbf{f}_0-\boldsymbol{g})+\lambda_0\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{f}_0$	Initial gradient
$\mathbf{p}_0:=-\mathbf{r}_0$	Initial search direction
$e_0:=\|\boldsymbol{r}_0\|^2$	
CG iterations	
$\mathbf{h}_n:=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\mathbf{p}_n+\lambda_n\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{p}_n$	Step search A
$\tau_n:=\mathbf{p}_n^{\mathrm{T}}\mathbf{r}_n/\mathbf{p}_n^{\mathrm{T}}\mathbf{h}_n$	Step search B
$\mathbf{f}_{n+1}:=\mathbf{f}_n+\tau_n\mathbf{p}_n$	HR image update
$\mathbf{r}_{n+1}^{C}:=\boldsymbol{C}\mathbf{f}_{n+1}-\boldsymbol{g}$	Gradient part update A
$\mathbf{r}_{n+1}^{R}:=\mathbf{R}\mathbf{f}_{n+1}$	Gradient part update B
$\lambda_{n+1}:=\mu\|\mathbf{r}_{n+1}^{C}\|_2/\|\mathbf{r}_{n+1}^{R}\|_2$	New λ
$\mathbf{r}_{n+1}:=\boldsymbol{C}^{\mathrm{T}}\mathbf{r}_{n+1}^{C}+\lambda_{n+1}\mathbf{R}^{\mathrm{T}}\mathbf{r}_{n+1}^{R}$	Final gradient update
$e_{n+1}:=\|\mathbf{r}_{n+1}\|^2$	
$\beta_n:=e_{n+1}/e_n$	
$\mathbf{p}_{n+1}:=-\mathbf{r}_{n+1}+\beta_n\mathbf{p}_n$	Search direction update
$n:=n+1$	
End CG iteration	

a specified tolerance, values lower than 10^{-6} are reasonable for images.

References

[1] R.Y. Tsai, T.S. Huang, Multiframe image restoration and registration, Advances Comput. Vision Image Process. 1 (1984) 317–339.

[2] S.P. Kim, N.K. Bose, H.M. Valenzuela, Recursive reconstruction of high resolution image from noisy undersampled multiframes, IEEE Trans. Acoust. Speech Signal Process. 38 (6) (1990) 1013–1027 [see also IEEE Trans. Signal Process.].

[3] H. Stark, P. Oskoui, High-resolution image recovery from image-plane arrays, using convex projections, J. Optical Soc. Amer. A 6 (1989) 1715–1726.

[4] A. Tekalp, M. Ozkan, M. Sezan, High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 1992, pp. 169–172.

[5] K. Aizawa, T. Komatsu, T. Saito, A scheme for acquiring very high-resolution images using multiplecameras, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992, vol. 3, 1992, pp. 289–292.

[6] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graph. Models Image Process. 53 (1991) 231–239.

[7] M. Irani, S. Peleg, Motion analysis for image enhancement: resolution, occlusion, and transparency, J. Visual Comm. Image Representation 4 (4) (1993) 324–335.

[8] R.R. Schultz, R.L. Stevenson, Improved definition video frame enhancement, in: International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, vol. 4, 1995, pp. 2169–2172.

[9] R.R. Schultz, R.L. Stevenson, Extraction of high-resolution frames from video sequences, IEEE Trans. Image Process. 5 (6) (1996) 996–1011.

[10] B.C. Tom, A.K. Katsaggelos, Resolution enhancement of video sequences using motion compensation, in: IEEE International Conference on Image Processing, vol. 1, 1996, pp. 713–716.

[11] R.C. Hardie, K.J. Barnard, E.E. Armstrong, Joint map registration and high-resolution image estimation using a sequence of undersampled images, IEEE Trans. Image Process. 6 (12) (1997) 1621–1633.

[12] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, in: SIAM Monographs on Matematical Modeling and Computation, SIAM, 1998.

[13] C.R. Vogel, Computational Methods for Inverse Problems, in: Frontiers in Applied Mathematics, SIAM, 2002.

[14] N. Nguyen, P. Milanfar, G. Golub, A computationally efficient super-resolution image reconstruction algorithm, IEEE Trans. Image Process. 10 (3) (2001) 573–583.

[15] N.K. Bose, S. Lertrattanapanich, J. Koo, Advances in super resolution using l-curve, in: the 2001 IEEE International Symposium on Circuits and Systems, vol. 2, 2001, pp. 433–436.

[16] T. Reginska, A regularization parameter in discrete ill-posed problems, SIAM J. Scientific Computing 17 (3) (1996) 740–749.

[17] M. Belge, M.E. Kilmer, E.L. Miller, Efficient determination of multiple regularization parameters in a generalized l-curve framework, Inverse Problems 18 (4) (2002) 1161–1183.

[18] M.G. Kang, A.K. Katsaggelos, General choice of the regularization functional in regularized image restoration, IEEE Trans. Image Process. 4 (5) (1995) 594–602.

[19] A. Bovik (Ed.), Handbook of Image and Video Processing, Academic Press, New York, 2000.

[20] N.P. Galatsanos, A.K. Katsaggelos, Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation, IEEE Trans. Image Process. 1 (3) (1992) 322–336.

[21] E.S. Lee, M.G. Kang, Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration, IEEE Trans. Image Process. 12 (7) (2003) 826–837.

[22] H. He, L.P. Kondi, An image super-resolution algorithm for different error levels per frame, IEEE Trans. Image Process. 15 (3) (2006) 592–603.

[23] M.-C. Hong, M.G. Kang, A.K. Katsaggelos, An iterative weighted regularized algorithm for improving the resolution of video sequences, in: IEEE International Conference on Image Processing, vol. 2, 1997, pp. 474–477.

[24] P.E. Eren, M.I. Sezan, A.M. Tekalp, Robust, object-based high-resolution image reconstruction from low-resolution video, IEEE Trans. Image Process. 6 (10) (1997) 1446–1451.

[25] R. Molina, M. Vega, J. Abad, A.K. Katsaggelos, Parameter estimation in Bayesian high-resolution image reconstruction with multisensors, IEEE Trans. Image Process. 12 (12) (2003) 1655–1667.

[26] R. Molina, A.K. Katsaggelos, J. Mateos, Bayesian and regularization methods for hyperparameter estimation in image restoration, IEEE Trans. Image Process. 8 (2) (1999) 231–246.

[27] S. Borman, R.L. Stevenson, Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors, in: IEEE International Conference on Image Processing, vol. 3, 1999, pp. 469–473.

[28] M.V.W. Zibetti, J. Mayer, Simultaneous super-resolution for video sequences, in: IEEE International Conference on Image Processing, vol. 1, 2005, pp. 877–880.

[29] M.V.W. Zibetti, J. Mayer, Outlier robust and edge-preserving simultaneous super-resolution, in: IEEE International Conference on Image Processing, vol. 1, 2006, pp. 1741–1744.

[30] M.V.W. Zibetti, J. Mayer, A robust and computationally efficient simultaneous super-resolution scheme for image sequences, IEEE Trans. Circuits and Systems for Video Technology 17 (10) (2007) 1288–1300.

[31] H.H. Barrett, K.J. Myers, Foundations of Image Science, in: Wiley Series in Pure and Applied Optics, Wiley, New York, 2004.

[32] A. Mohammad-Djafari, Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems, in: IEEE International Conference on Image Processing, 1996, pp. 473–472.

[33] S. Chaudhuri, Super-Resolution Imaging, in: Kluwer International Series in Engineering and Computer Science, Kluwer, Dordrecht, 2001.

[34] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Process. Magazine 20 (3) (2003) 21–36.

[35] S. Borman, R. Stevenson, Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying psf, in: SPIE Electronic Imaging, 2004.

[36] C. Stiller, J. Konrad, Estimating motion in image sequences, IEEE Signal Process. Magazine 16 (4) (1999) 70–91.

[37] R.R. Schultz, L. Meng, R.L. Stevenson, Subpixel motion estimation for super-resolution image sequence enhancement, J. Visual Comm. Image Representation 9 (1) (1998) 38–50.

[38] S. Borman, R.L. Stevenson, Super-resolution from image sequences—a review, in: Proceedings of 1998 Midwest Symposium on Circuits and Systems, 1999, pp. 374–378.

[39] A. Mohammad-Djafari, A full Bayesian approach for inverse problems, in: 15th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt95), 1995.

[40] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1991.

[41] D.G. Luenberger, Introduction to Linear and Nonlinear Programming, Addisson Wesley, Reading, MA, 1973.

[42] F.S.V. Bazán, Fixed-point iterations in determining the tikhonov regularization parameter, Inverse Problems 24(3).

[43] M.V.W. Zibetti, F.S.V. Bazan, J. Mayer, Determining the parameters in regularized super-resolution reconstruction, in: IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, 2008, pp. 853–856.

[44] A.E. Gamal, H. Eltoukhy, Cmos image sensors, IEEE Circuits Devices Magazine 21 (3) (2005) 6–20.

[45] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intell. 17 (1981) 185–203.

[46] D. Hasler, L. Sbaiz, S. Süsstrunk, M. Vetterli, Outlier modeling in image matching, IEEE Trans. Pattern Anal. Machine Intell. 25 (3) (2003) 301–315.

[47] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super-resolution, IEEE Trans. Image Process. 13 (10) (2004) 1327–1344.