On the amount of regularization for super-resolution reconstruction

Yann Traonmilin, Saïd Ladjal, Andrés Almansa

On the amount of regularization for super-resolution reconstruction

Yann Traonmilin*, Saïd Ladjal, Andrés Almansa

***Abstract*—Modern digital cameras are quickly reaching the fundamental physical limit of their native resolution. Super-resolution (SR) aims at overcoming this limit. SR combines several images of the same scene into a high resolution image by using differences in sampling caused by camera motion. The main difficulty encountered when designing SR algorithms is that the general SR problem is ill-posed. Assumptions on the regularity of the image are then needed to perform SR. Thanks to advances in regularization priors for natural images, producing visually plausible images becomes possible. However, regularization may cause a loss of details. Therefore, we argue that regularization should be used as sparingly as possible, especially when the restored image is needed for further precise processing. This paper provides principles guiding the local choice of regularization parameters for SR. With this aim, we give an invertibility condition for affine SR interpolation. When this condition holds, we study the conditioning of the interpolation and affine motion estimation problems. We show that these problems are more likely to be well posed for a large number of images. When conditioning is bad, we propose a local total variation regularization for interpolation and show its application to multi-image demosaicking.**

EDICS : TEC-ISR Interpolation, Super-Resolution and Mosaicing: Interpolation and superresolution; Mosaicing, registration and alignment; Multi-image fusion

I. Introduction

A. Problem statement and state of the art

Super-resolution aims at recovering a high resolution (HR) image from several low resolution (LR) images. In the most generic formulation of SR, we need to estimate camera blur, motion and the HR image simultaneously, which is an ill-posed problem. SR techniques have been reviewed several times in the literature [1], [2]. They mostly rely on a regularized minimization of a functional linking the acquired LR images and the unknown HR image. There is a wide choice of such functionals, including L^2 norm with Tychonov regularization [3], bilateral total variation (TV) regularization [4], [5], and L^1 norm with TV regularization [6], [7]. The choice of a regularizer is an implicit hypothesis (or a priori information) on the content of the image. For example, perfect reconstruction with TV regularization is not possible if the image contains too many textures [8], [9]. Even recent non-local regularization methods for single and multi-image super-resolution [10], [11] need images which exhibit rotational [12] or multiscale [13] self-similarities. Our aim is to avoid or minimize assumptions made on the HR image, and consequently minimize the amount of regularization.

The relative motion between LR images is often restricted to translations and rotations. However, a small motion of the camera in the depth direction can cause a linear zoom between images. To describe this, we will consider affine motions of the LR sampling grids. Affine motions have been considered in [14]–[16] where conventional techniques for parameter estimation and regularized reconstruction are described. As we study the well-posedness of the SR problem, we will not try to invert camera blur as this part of the SR problem is generally ill-posed. With this configuration, super-resolution can be split in two processes: *(i)* interpolation (*i.e.* SR inversion with known motion parameters) and *(ii)* motion parameter estimation. In the interpolation case, super-resolution is an irregular to regular sampling interpolation problem which could be solved with general techniques [17]–[20]. With super-resolution sampling configurations, we can use the fact that each LR image is acquired on a regular grid to obtain dedicated results and methods. For example, the pure translational case can be viewed as a multichannel sampling problem. Thus, extending the result of Papoulis [21], Ahuja and Bose [22] showed that if we have a super-resolution factor M, only M^2 LR images with pure translational motions are needed to perfectly recover the HR image in a noiseless set-up. In this case, no assumption on the content of the HR image is needed (apart from the fact that it is band-limited). It naturally leads to the following question: in what conditions on the acquisition system can we perform super-resolution without hypotheses on the image?

The other side of the SR reconstruction problem is the registration of each LR image. It consists in estimating motion parameters between LR images (6 parameters for each affine motion). Some techniques for affine motion estimation between a pair of well-sampled images already exist [23]. However, in a non-trivial super-resolution set-up, LR images are assumed to be aliased, and these methods do not give estimates precise enough to perform a good reconstruction [24], [25]. If the interpolation problem is invertible, SR reconstruction can be viewed as a non-linear minimization problem with respect to motion parameters which is called variable projection [26]. [27] showed that using a regularized variable projection method for super-resolution reconstruction gives stable results in practice. However, the necessity for a regularization term in the variable projection method is not questioned.

When the acquired data is contaminated by noise, having a

Y. Traonmilin, S. Ladjal and A. Almansa (email: {yann.traonmilin,ladjal,andres.almansa}@telecom-paristech.fr) are with Telecom ParisTech, CNRS LTCI, 46 rue Barrault, 75013 Paris, France. Tel: +33145817777

Work partially funded by FUI-9 CEDCA project.

good conditioning of the system is critical for the quality of the reconstructed image. The influence of the SR zoom has been studied in the case of interpolation in [28]. Experiments on the Cramer-Rao bound with respect to the number of images were shown by Robinson [25] and Champagnat [29], demonstrating that the reconstruction error of the pseudo-inverse decreases when the number of images grows.

B. Overview and Contributions

In this paper, we first describe (**Section II**) the theoretical context of super-resolution with affine motion, which requires special attention in terms of how hypotheses are formulated, and under which conditions the super-resolution, denoising and deconvolution problems can be decoupled from one another. Within this framework we give in **Section III** a sufficient condition on the invertibility of SR interpolation. This extends the work of Ahuja [22] (which is restricted to translational motions) to the invertibility of affine motion SR interpolation. This result allows to show that the number of samples needed for affine SR with random motions is the same as the one for random sampling of trigonometric polynomials as shown by Bass & Gröchenig [30]. In **Section IV**, the asymptotic behaviour of translational SR interpolation shown by Champagnat [29] is extended to affine motions. We show that it is in the interest of acquisition system designers to target the acquisition of a large number of LR images, as the SR reconstruction is likely to be well-posed in that setting. We also study the registration for large number of images by calculating the Hessian of the non-linear least squares motion parameter estimation problem. It allows for a better understanding of how variable projection behaves and shows that a regularization term (like in [27]) is not always necessary. In **Section V**, we show that the conditioning is spatially varying. We calculate this conditioning for small motions and use it to predict how SR interpolation must be regularized. This prediction takes the form of a new local weighting scheme of the total variation regularizer. Finally, we apply it to multi-image demosaicking. We show that regularization and cross-channel dependencies used by even the most advanced demosaicking algorithms can be avoided if multiple images are available. In this case regularization assumptions are minimized by our procedure, and demosaicking artifacts are avoided.

II. Theoretical considerations

A. Problem set-up

The purpose of the super-resolution problem is to invert a linear map A that produces N LR images from a single HR image:

$$\begin{aligned} A : (\mathbb{C}^{ML\times ML}) &\rightarrow (\mathbb{C}^{L\times L})^N \\ u &\rightarrow (SQ_i u)_{i=1,N} \end{aligned} \tag{1}$$

where N is the number of LR images, M is the super resolution factor, $L \times L$ is the size of a LR image, u is the HR image. S is the sub-sampling operator by a factor M and

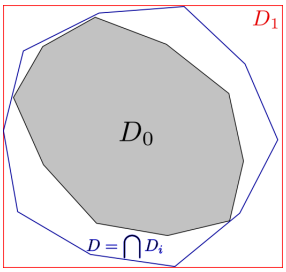


Fig. 1. Representation of the signal spectrum and the spectral Nyquist limits of the HR grids with affinities applied

Q_i are the affine deformation associated with each LR image. We call q_i the corresponding affine motion of the sampling grid *i.e.*

$$(Q_i u)(\mathbf{x}) = u(q_i(\mathbf{x})). \tag{2}$$

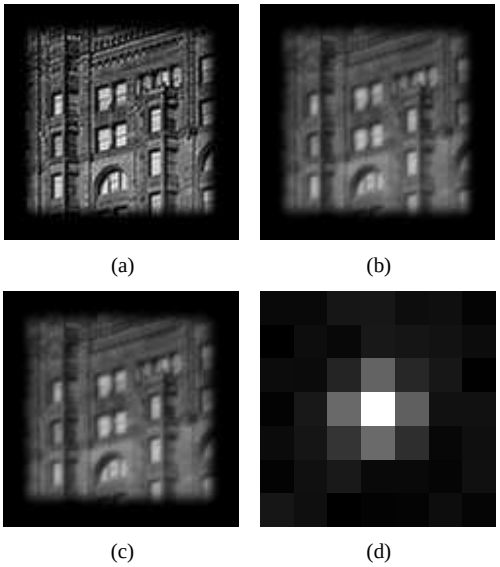
We decompose q_i in its linear and translational parts: $q_i\mathbf{x} = l_i\mathbf{x} + t_i$. l_i are 2×2 matrices, t_i are 2 dimensional vectors. SR is the process of recovering u from $w = Au + n$ (n is acquisition noise). If the Q_i are known, the inversion of A is called the super-resolution interpolation. The recovery of both u and Q_i is called super-resolution reconstruction.

B. Hypotheses

We need to make some assumptions on the problem to simplify the study:

1) u is band-limited
2) the first LR image is the reference image ($Q_1 = Id$)
3) affine maps on the coordinates (q_i) are invertible
4) affine maps do not generate aliasing on the HR grid

Condition 4) deserves further explanation: Let us consider a continuous image formation model. LR images are generated by $w_i = SS_{HR}Q_i u$, where S_{HR} is the sampling with the HR sampling step. In a practical SR algorithm, we try to estimate $S_{HR}u$, *i.e.* we commute the HR sampling and the motion. It is possible to do so if u is well sampled by S_{HR} before and after the motion. We show that this is equivalent to an adequate choice of the HR sampling step.
Let D_0 be the support of the spectrum of u. Condition 4) means that each $S_{HR}Q_i u$ image is not aliased. For each affine motion, the frequency domain satisfying the Shannon condition is given by its reciprocal cell D_i. u is well sampled after any motion if its spectrum is contained in $D = \bigcap D_i$.More over D_1 is a square containing D. We set the HR sampling step such that $D_0 \subset D$. The configuration of these frequency domains is shown in Figure 1. In other words, from a given band-limited signal, we can always define a HR sampling step which does not cause aliasing of the continuous reference image u after motion. The strength of this assumption depends on the amplitude of the affine maps. The smaller they are, the weaker it is. As our images are spatially limited, we only consider the reconstruction of the trigonometric polynomials associated with them.



(a) (b) (c) (d)

Fig. 2. Validity of the assumption of commutativity: (a) HR image (b) Blurred HR image (c) image reconstructed from 16 LR images with $M = 2$ by neglecting the blur. PSNR calculated with respect to (b) is 46.7 (d) blur kernel

C. About camera blur

Camera blur (noted F) happens just before sub-sampling:

$$w_i = SFQ_i u \tag{3}$$

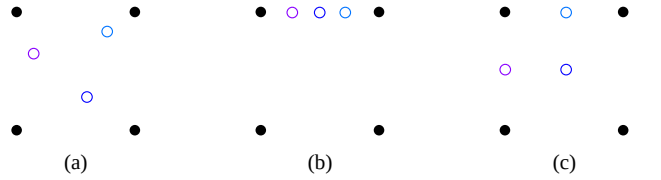
In this paper, we suppose that it is identical for all LR images and that we can commute blur and motion. This assumption holds for purely translational motions, and for rotations (as long as the blur kernel is rotationally symmetric). The supposition is also justified for more general blur kernels in the case of small motions. In fact, the difference between the least-squares estimate $\tilde{u}$, when we do not take into account camera blur and the filtered HR image Fu is:

$$\begin{aligned} \tilde{u} - Fu &= (A^H A)^{-1} \left(\sum_i Q_i S^H w_i - (A^H A) Fu \right) \\ &= (A^H A)^{-1} \left(\sum_i Q_i S^H S (FQ_i - Q_i F) u \right) \end{aligned} \tag{4}$$

This assumption is valid if the energy of $S(FQ_i - Q_iF)u$ is smaller than the energy of the acquisition noise. We show an experiment of reconstruction with small affinities in Figure 2: by neglecting the blur in the model, we reconstruct the blurred version of the HR image.

III. Invertibility condition of the SR interpolation

If the problem of SR interpolation is not invertible, the SR problem is ill-posed, and regularization will play the most important part in the interpolation process. Interpolating without regularization requires an invertible problem. As our objective is the study of cases where no regularization is needed, we begin by studying the critical condition for invertibility in terms of the number of images. We will then be able to distinguish two invertible cases: well conditioned and badly conditioned (which will be discussed in the following sections). We give a sufficient condition (on the motion of the sampling grids) for the invertibility with the hypotheses defined in the previous section.



(a) (b) (c)

Fig. 3. Invertible and non invertible configurations covered by Theorem III.1 for $M = 2$ and pure translations. Black dots represent the reference grid (a) Invertible case meeting hypotheses (b) non invertible case excluded by the theorem (c) configuration excluded by the theorem but still invertible

Let us name the following sampling grids: $\Gamma^{hr} = [1, ML]^2 \subset \mathbb{Z}^2$ and $\Gamma = M.[1, L]^2$. Γ^c is the complement of Γ in Γ^{hr}, i.e. the support of images in the kernel of S. We give a sufficient condition for A to be invertible: the difference between the motion of two positions in Γ^c must not be an integer, coordinate by coordinate.

Theorem III.1. *If $N \geq M^2$ and for all $p_i, p_j \in \Gamma^c, 1 \leq k_1 < k_2 \leq N, ||q_{k_1}^{-1} p_i - q_{k_2}^{-1} p_j \mod 1||_0 = 2$, A is injective.*

Proof: We show by induction over N that when adding a LR image, the dimension of the kernel of the function A decreases by a factor L^2. For clarity, the proofs of the necessary lemmas are shown in the appendix. Let:

$$\begin{aligned} A_n : (\mathbb{C}^{ML \times ML}) &\rightarrow (\mathbb{C}^{L \times L}) \\ u &\rightarrow SQ_n u \end{aligned} \tag{5}$$

We prove : for all $1 < n \leq M^2$, $\dim \cap_{k=1,n} \ker Q_k = (M^2 - n)L^2$

For $n = 2$: let $p_i \in \Gamma^c$. Let $v_i = \mathbf{1}_{p_i}$. Let $u_i = Q_1^{-1} v_i$. We have $Sv_i = 0$. Consequently $A_1 u_i = 0$ and $u_i \in \ker A_1$. We just defined $(M^2 - 1)L^2$ independent u_i generating $\ker A_1$: $\text{span}(u_i)_{i=1,(M^2-1)L^2} = \ker A_1$. Similarly we construct $\text{span}(u_i')_{i=1,(M^2-1)L^2} = \ker A_2$. With Lemma VII.2 ($ker A_1 + \ker A_2 = \mathbb{C}^{ML \times ML}$), the dimension of the intersection is:

$$\begin{aligned} \dim(\ker A_1 \cap \ker A_2) =& \dim(\ker A_1) + \dim(\ker A_2) \\ & - \dim(\ker A_1 + \ker A_2) \\ =& (M^2 - 2)L^2 \end{aligned} \tag{6}$$

Let $n > 2$. Let us suppose that $\dim \cap_{k=1,n} \ker A_k = (M^2 - n)L^2$. We use Lemma VII.3: $(\cap_{k=1,n} \ker A_k) + \ker A_{n+1} = \mathbb{C}^{ML \times ML}$. By using the same dimensions relation as for $n = 2$, we get the result for $n + 1$. ■

The main condition is the number of images which is the same as the condition for translations by Papoulis. Example of non invertible and invertible configurations are shown in Figure 3. A direct corollary is that, if the motion parameters are random, the SR interpolation problem is almost-surely invertible:

Corollary III.1. *If $N \geq M^2$ and motion parameters are random, A is injective almost-surely.*

Proof: The space of excluded affine motion parameters $E \subset \mathbb{R}^{6N}$ in Theorem III.1 has measure 0. Let $\theta \in \mathbb{R}^{6N}$ be a parameter vector. $(\mathbb{R}.\theta) \cap E$ has measure 0 because it is countable. By using Fubini, E has measure 0 in $\mathbb{R}^6$. ■

We can compare this result with [30] where it was shown that the problem of random sampling of trigonometric polynomials is invertible almost surely if there are at least as many equations as unknowns (equivalent to $N = M^2$). Our probabilistic result is different because sampling locations are not completely random. We also gave a deterministic condition for this invertibility. This condition excludes non invertible cases. For example, if the motion is translational in only one direction, the hypothesis of the theorem is not met (as seen in Figure 3).

IV. Conditioning of the SR problem with respect to the number of images

A. Interpolation

Measuring the difficulty of a linear inversion problem, such as super resolution interpolation is often made using the condition number of the linear map defining the problem. Several factors have an influence on this conditioning. Baker [28], shows that the condition number of the system grows with the super-resolution factor M. The condition number also depends on the sampling distribution [17]. For a controlled motion, a condition number of 1 is obtained by regularly spacing LR grids matching the HR grid when merged. When motions are random and uniform, the condition number of the affine SR interpolation problem converges to one when the number of images grows. This fact was experimentally illustrated for translational SR in [25] in terms of the Cramer-Rao bound for HR image estimation. [29] shows that the reconstruction error decreases to 0 when the number of images grows. We suppose that $N \geq M^2$ and that the affine motions respect the condition of Theorem III.1. In this case, the HR image can be perfectly recovered by taking the pseudo inverse of A. The condition number of the system is the ratio of the extremal eigenvalues of $A^H A$. We show the following using a similar technique as in [29]:

Proposition IV.1. *Let us suppose that the affine motions q_i have the following distribution: t_i are uniform in $[0, M]^2$ and the average of the l_i is Id. Then the conditioning κ of the system converges to 1 (in the distribution sense) as the number of images grows.*

Proof: To recover u at a particular pulsation $\omega \in \mathbb{R}^2$, we first write $\hat{u}(\omega)$ as a linear combination of the $\hat{w}_i(\omega)$:

$$\hat{w}_i(l_i^{-1}\omega) = \frac{1}{M^2} \sum_{k \in \mathbb{Z}^2} \hat{u}(\omega + l_i \frac{2\pi k}{M}) e^{j(\omega.t_i + \frac{2\pi k}{M}.(l_i^{-1}t_i))} \quad (7)$$

Only M^2 terms in the sum are non-zero. If there is more than M LR images, it is an overdetermined system of size $N \times M^2$ for each pulsation:

$$C\hat{u}_{al} = w \quad (8)$$

where $\hat{u}_{al}(k) = \hat{u}(\omega + l_i \frac{2\pi k}{M})$, $w(i) = \hat{w}_i(l_i^{-1}\omega)$ and $C_{i,k} = \frac{1}{M} e^{j(\omega.t_i + \frac{2\pi k}{M}.(l_i^{-1}t_i))}$. C is the product of 2 matrices:

$$C = \Delta B \quad (9)$$

with $\Delta = \frac{1}{M^2} \operatorname{diag}(e^{j\omega.t_i})$ and $B_{i,k} = e^{j\frac{2\pi k}{M}.(l_i^{-1}t_i)}$. The conditioning of the system is consequently the conditioning of $R = B^H B$ which is a Toeplitz matrix with term $R_{r,s} = \sum_i e^{j\frac{2\pi(s-r)}{M}.(l_i^{-1}t_i)}$. We can show with a direct application of the central limit theorem that R converges to a multiple of identity because the complex numbers $e^{j\frac{2\pi(s-r)}{M}.(l_i^{-1}t_i)}$ converge to a uniform distribution on the unit circle (for $s \neq r$). By continuity of the condition number, the condition number $\kappa(R)$ converges to 1. ■

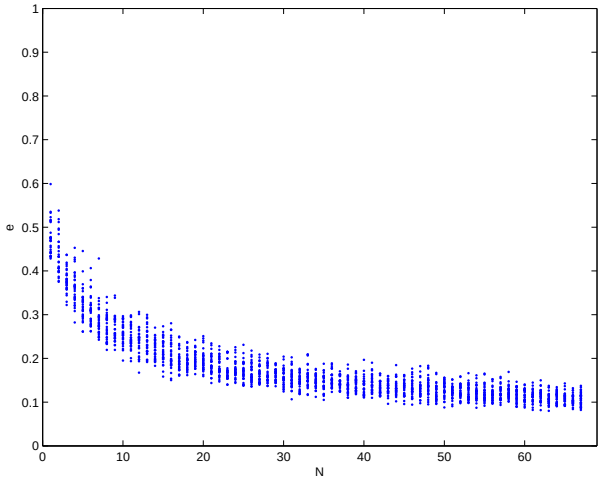


Fig. 4. Convergence of the estimator. Interpolation error with respect to the number of LR images

A large number of images is statistically better. This result can be illustrated by the following reconstruction method: as $\kappa(R) \to 1$, $R \sim N.Id$ and $A^\dagger \sim \frac{1}{N} A^H$. We can use $\frac{1}{N} A^H$ as a reconstruction operator for a large number of images. We plot in Figure 4 the reconstruction error $e_N = ||u - u_N||$ of $u_N = \frac{1}{N} A^H w$ with respect to N. For each N value, we generated 30 experiments with random affine motion distributed as in Proposition IV.1. The same HR (from Figure 2) image is used for all experiments.

We now show that it is always interesting to add an observation (a LR image). The reconstruction noise decreases with the number of images. Our observation model is:

$$w = Au + n \quad (10)$$

$n \sim N(0, \sigma^2 I)$ is a white zero mean Gaussian noise. The reconstruction noise will be:

$$n_r = A^\dagger n \quad (11)$$

and $n_r \sim N(0, \sigma^2 A^\dagger A^{\dagger H}) = N(0, \sigma^2 (A^H A)^{-1})$. The reconstruction noise will have normalized energy $e = \sigma^2 tr((A^H A)^{-1}) = \sum_i \frac{1}{\lambda_i}$ where λ_i are the eigenvalues (e.v.) of $A^H A$.

We consider adding an image:

$$\begin{aligned} A' : (\mathbb{C}^{ML \times ML}) &\to (\mathbb{C}^{L \times L})^{N+1} \\ u &\to (SQ_k u)_{k=1,N+1} \end{aligned} \quad (12)$$

$$w' = A'u + n' \tag{13}$$

with $n' \sim N(0, \sigma^2 I)$. We have : $e' = \sigma^2 tr((A'^H A')^{-1}) = \sum \frac{1}{\lambda'_i}$ with $\lambda'_i = e.v.(A'^H A')$.

Proposition IV.2. *Acquiring more images diminishes the noise, i.e.* $e' < e$

Proof: We first prove that $\lambda'_i \geq \lambda_i$ for all i. Using Weyl's inequalities:

$$\lambda_i \leq \lambda'_i + \lambda_{max}(A^H A - A'^H A') \tag{14}$$

We have:

$$\begin{aligned} A'^H A' - A^H A &= \sum_{k=1,N+1} Q_k^H S^H S Q_k - \sum_{k=1,N} Q_k^H S^H S Q_k \\ &= Q_{N+1}^H S^H S Q_{N+1} \end{aligned} \tag{15}$$

which is a positive linear map. Thus $\lambda_{max}(A^H A - A'^H A') \leq 0$ and $\lambda_i \leq \lambda'_i$. With this result we have $\frac{1}{\lambda'_i} \leq \frac{1}{\lambda_i}$ for all i. Consequently:

$$\begin{aligned} \sum_i \frac{1}{\lambda'_i} &\leq \sum_i \frac{1}{\lambda_i} \\ \sigma^2 \sum_i \frac{1}{\lambda'_i} &\leq \sigma^2 \sum_i \frac{1}{\lambda_i} \\ e' &\leq e \end{aligned} \tag{16}$$

■

This result shows that if we are able to recover the motion parameters, the best strategy is to keep all the available LR images for the reconstruction. This result matches the intuitive idea that having more data points increases the signal-to-noise ratio. Still, it must be noted that this property is dependent on the structure of the linear map generating the data. We discuss in the next section the difficulty of recovering the motion parameters.

B. Parameter estimation

In a noiseless case, when $N > M^2$, $u = (A^H A)^{-1} A w = A^\dagger(\boldsymbol{\theta}_0) w$. $\boldsymbol{\theta}_0$ are the parameters of affine motions. To estimate $\boldsymbol{\theta}_0$, we try to minimize [5], [26]:

$$G(\tilde{\boldsymbol{\theta}}) = ||A(\tilde{\boldsymbol{\theta}}) A^\dagger(\tilde{\boldsymbol{\theta}}) w - w||_2^2 \tag{17}$$

which is not a convex problem. With a first estimate (which can be obtained with a dedicated registration technique), we can minimize this functional with a gradient descent. The speed of convergence and the precision of this method will depend directly on the conditioning of the Hessian H at $\boldsymbol{\theta}_0$ (we suppose that the first LR image is not translated, and that H is not singular). We calculate this condition number. The gradient has the following expression:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} G(\boldsymbol{\theta}) &= 2(A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta}) w - w)^H (\frac{\partial}{\partial \theta_i}[A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta})] w) \\ &= 2(A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta}) w - w)^H (\frac{\partial}{\partial \theta_i}[A(\boldsymbol{\theta})] A^\dagger(\boldsymbol{\theta}) w \\ &+ A(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_i}[A^\dagger(\boldsymbol{\theta})] w) \\ &= 2(A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta}) w - w)^H (\frac{\partial}{\partial \theta_i}[A(\boldsymbol{\theta})] A^\dagger(\boldsymbol{\theta})) \end{aligned} \tag{18}$$

where the last line was obtained by orthogonality. The Hessian is then (we do not calculate the constant):

$$\begin{aligned} H_{i,j} \propto &(\frac{\partial}{\partial \theta_j}[A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta})] w)^H (\frac{\partial}{\partial \theta_i}[A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta})] w) \\ &+ (A(\boldsymbol{\theta}) A^\dagger(\boldsymbol{\theta}) w - w)^H (\frac{\partial^2}{\partial \theta_j \partial \theta_i}[A(\boldsymbol{\theta})] A^\dagger(\boldsymbol{\theta}) w) \end{aligned} \tag{19}$$

In particular, the Hessian at $\boldsymbol{\theta}_0$ is:

$$H_{i,j} \propto (\frac{\partial}{\partial \theta_j}[A(\boldsymbol{\theta}_0) A^\dagger(\boldsymbol{\theta}_0)] w)^H (\frac{\partial}{\partial \theta_i}[A(\boldsymbol{\theta}_0)] A^\dagger(\boldsymbol{\theta}_0) w) \tag{20}$$

We set $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (we write $A = A(\boldsymbol{\theta}_0)$), as our aim is to calculate the conditioning of the Hessian at the minimum of the functional. Calculations (developed in the annex) lead to the following expression of the Hessian:

$$H_{i,j} \propto \left\langle \frac{\partial}{\partial \theta_j}[A] u \,\middle|\, (I - AA^\dagger) \frac{\partial}{\partial \theta_i}[A] u \right\rangle \tag{21}$$

We call $\frac{\partial}{\partial \theta_j}[A] u = w'_j$:

$$H_{i,j} \propto \langle w'_j,\, (I - AA^\dagger) w'_i \rangle \tag{22}$$

Because LR images are separated:

$$H_{i,j} \propto \gamma_{i,j} \langle w'_j,\, w'_i \rangle - \langle A^H w'_j,\, A^\dagger w'_i \rangle \tag{23}$$

where $\gamma_{i,j} = 1$ when θ_i and θ_j are parameters related to the motion of the same LR image, 0 otherwise. $\gamma_{i,j} < w'_j, w'_i >$ define a block diagonal symmetrical matrix H_B. Furthermore, all blocks are of size 6×6 and bounded in the 2-norm sense. From the previous part, we have $A^\dagger \sim \frac{1}{N} A^H$. Consequently, with Cauchy Schwartz inequality:

$$\begin{aligned} |\langle A^H w'_j,\, A^\dagger w'_i \rangle| &= O(\frac{1}{N}) |\langle A^H w'_j,\, A^H w'_i \rangle| \\ &= O(\frac{1}{N}) ||A^H w'_j|| ||A^H w'_i|| \end{aligned} \tag{24}$$

As the quantities $||A^H w'_j||$ are bounded:

$$|\langle A^H w'_j,\, A^\dagger w'_i \rangle| = O(\frac{1}{N}) \tag{25}$$

and

$$H_{i,j} \propto \gamma_{i,j} \langle w'_j,\, w'_i \rangle (1 - O((\frac{1}{N})) \sim_{N \to \infty} \gamma_{i,j} \langle w'_j,\, w'_i \rangle \tag{26}$$

we finally have : $H \to H_0$ and $\kappa(H) \to \kappa(H_B)$ by continuity of the condition number. Let us calculate one 6×6 block H_B corresponding to one LR image k_0. We suppose that θ_j and θ_i

N	4	12	40
$\|\|\tilde{u}_0 - u\|\|/\|\|u\|\|$	3.19	0.2	0.1
$\|\|\tilde{u} - u\|\|/\|\|u\|\|$	2.5	0.06	0.04

TABLE I
AVERAGE RECONSTRUCTION ERROR WITH RESPECT TO THE NUMBER OF IMAGES

are two motion parameters of the motion $Q = Q_{k_0}$. We have:

$$\begin{aligned} H_{Bi,j} &\propto \left\langle \frac{\partial}{\partial\theta_j}[A]u \,\middle|\, \frac{\partial}{\partial\theta_i}[A]u \right\rangle \\ &\propto \left\langle S\frac{\partial}{\partial\theta_j}[Q]u \,\middle|\, S\frac{\partial}{\partial\theta_i}[Q]u \right\rangle \end{aligned} \tag{27}$$

H_B represent the conditioning relative to the estimation of parameters of one affine transformation. We calculated experimentally eigenvalues of 1000 different H_B matrices using equation (27). We generated u as a Gaussian i.i.d. process to have a full band signal which is uncorrelated with the acquisition parameters. It leads to a maximum possible condition number for H_0 under 100. Compared to a pure 1D translational case where H_0 is diagonal with condition number 1, the more general affine case is slightly more difficult due to the intrinsic difficulty of motion estimation. In practice, this difficulty is limited as shown by the next experiment.
In Table I, we calculated the average reconstruction error of 10 experiments with parameter estimation with respect to the number of images. We used a non linear conjugate gradient algorithm, stopped after the same number of iterations for each number of images. For each experiment, the starting point of the non linear conjugate gradient algorithm $\tilde{u}_0$ is a random perturbation of the solution (simulation of an estimation with a LR method).

C. Summary

In this section, we studied the evolution of the conditioning with N. For SR reconstruction, interpolation and parameter estimation each play a part in the difficulty of the problem. When we make the hypothesis that motions are random with a reasonable distribution, the interpolation part converges to a conditioning of 1 when N grows. The global estimation part is more difficult as each affine motion has an intrinsic conditioning for its estimation. This conditioning depends on the frequency content of the HR image and on the value of the motion parameter. If the motion is a translation, this conditioning is 1 for full band signals. In the generic affine case, when the number of images grows and no motion is degenerate (the zoom part of affinities is close to identity), the conditioning becomes good experimentally.

V. LOCAL CONDITIONING AND REGULARIZATION

A. Local conditioning

The fusion of LR grids is a sampling grid which is generally not periodic. If motions are small, we observe local variations of the spatial distribution of the samples leading to a spatial variability in the noise generated in the inversion process. In this section, we predict this conditioning in the case of small motions (which is a reasonable hypothesis for hand held camera), and use this prediction to adaptively regularize with respect to local conditioning.
We study the conditioning in the critical case $N \geq M^2$ where the problem is invertible (from the previous section). We propose to use the conditioning of an equivalent pure translational SR problem at each location. It is justified by the comparison of the reconstruction noise n_{rec} of the system and the reconstruction noise of a pure translational model. When the LR images are contaminated by a noise $\mathbf{n}$:

$$n_{rec} = A^{\dagger}\mathbf{n} \tag{28}$$

We calculate the power of the noise locally. We restrict the image space of the application $A^{\dagger}$ to one LR pixel in the HR image space to study its local behavior. Let $\mathbf{x}_0 = [x_0, y_0]$. Let $\mathbf{x} \in [x_0, x_0 + M - 1] \times [y_0, y_0 + M - 1] = D \subset \mathbb{Z}^2$. Let $\mathbf{1}_{\mathbf{x}}$ be the indicator function $\mathbf{x} \in D$ in the HR image. We now consider the mapping:

$$\begin{aligned} A^{\dagger}_{\mathbf{x}_0} : E = A(\mathrm{span}((\mathbf{1}_{\mathbf{x}})_{\mathbf{x}\in D}) &\to F = \mathrm{span}((\mathbf{1}_{\mathbf{x}})_{\mathbf{x}\in D}) \\ w &\to A^{\dagger}w \end{aligned} \tag{29}$$

We call local conditioning at position $\mathbf{x}_0$, the conditioning of $A^{\dagger}_{\mathbf{x}_0}$. This conditioning is the ratio of the bounds of the quantity (greatest and smallest singular values):

$$||A^{\dagger}_{\mathbf{x}_0}w||, ||w|| = 1 \tag{30}$$

We can calculate equivalently the bounds of $||A_{\mathbf{x}_0}u||, ||u|| = 1$. Let $u = \sum b_k \mathbf{1}_{\mathbf{x_k}} \in F$ with $||u|| = 1$. We have :

$$\begin{aligned} ||A_{\mathbf{x}_0}u||^2 &= ||\sum\ b_k A\mathbf{1}_{\mathbf{x_k}}||^2 \\ &= \sum_{k_1,k_2} \bar{b}_{k_1} b_{k_2} (\mathbf{1}_{\mathbf{x_{k_1}}})^H A^H A \mathbf{1}_{\mathbf{x_{k_2}}} \\ = \sum_{k_1,k_2} \bar{b}_{k_1} b_{k_2} \sum_{i=1,N} \sum_{\mathbf{y}\in\Gamma} &\mathrm{sincd}(\mathbf{y} - \tau_{i,k_1})\mathrm{sincd}(\mathbf{y} - \tau_{i,k_2}) \end{aligned} \tag{31}$$

where sincd is the finite discrete Shannon interpolator and $\tau_{i,k} = q_i\mathbf{x_k}$. Because sincd is differentiable, we can use the mean value theorem to compare this expression to a pure translational one and obtain an expression of the form:

$$\left| \|A_{\mathbf{x}_0}u\|^2 - \|A^{tr}_{\mathbf{x}_0}u\|^2 \right| \leq K\|\boldsymbol{\theta} - \boldsymbol{\theta}^{tr}\|^2 \tag{32}$$

where $\boldsymbol{\theta} = (q_i)_i$ is the set of translations induced by the motion, $\boldsymbol{\theta}^{tr}$ is $\boldsymbol{\theta}$ averaged over the HR pixel (over index k) and $A^{tr}_{\mathbf{x}_0}$ is the pure translational SR operator associated with $\boldsymbol{\theta}^{tr}$ and K is a constant which does not depend on $\mathbf{x}_0$. Thus, for sufficiently small motions, the noise of the system will behave as in a pure translational case. Experiments showed that for affinities in a small range (rotation in the range -5,+5 degrees, zoom in the range $\times 0.9, \times 1.1$), we can use $\kappa(\mathbf{x}_0) = \mathrm{cond}(R)$ as a local conditioning measure, with R defined as in Part IV with the translations $\boldsymbol{\theta}^{tr}$. The result of the this prediction if shown in Figure 5.

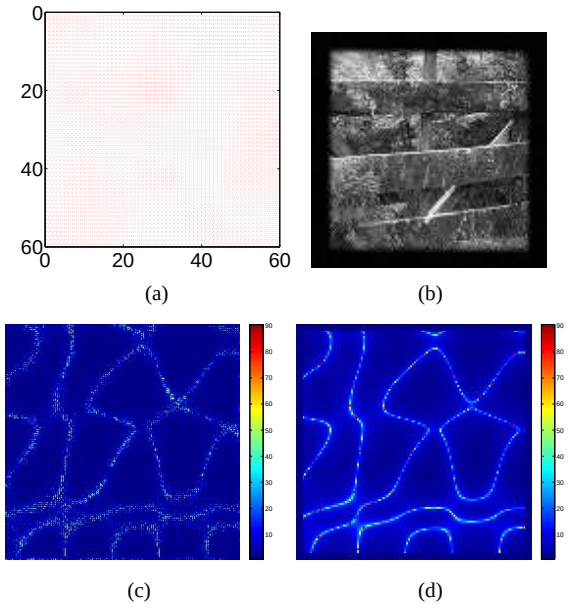


Fig. 5. Local conditioning of the SR problem. (a) Zoom on the fusion of the 4 LR grids (60×60 pixels upper left corner). (b) Example of a LR image. (c) Amplitude of the reconstruction noise (sampled on a LR grid) normalized by the input noise variance. (d) Local conditioning $2\sqrt{\kappa(\mathbf{x})}$.

B. Local regularization

Two types of regularization have been used mostly: Tychonov regularization and TV (or bilateral TV) regularization [1], [3], [7], [25]. We propose a local total variation regularization scheme where our local conditioning measure defines weights for the total variation term. In [31] the authors proposed to weight the bilateral TV by the diagonal entries of the operator which is an empirical way of taking into account local conditioning. We minimize the function:

$$\begin{aligned} J_\alpha(\tilde{u}) &= G(\tilde{u}) + \lambda H_\alpha(\tilde{u}) \\ H_\alpha(\tilde{u}) &= \int \alpha . |\nabla \tilde{u}| \end{aligned} \tag{33}$$

where $\alpha(\mathbf{x}) = \sqrt{1 - \frac{1}{\kappa(\mathbf{x})}}$ and λ is the regularization parameter. When $\alpha = 1$, H is a conventional total variation regularizer. In [32], $\alpha(\mathbf{x}) = log(\kappa(\mathbf{x}))$ was chosen experimentally. We now show the new choice is optimal.

Let us consider a pure translational SR problem with global TV regularization. We suppose that optimal TV regularization gives a good approximation of U. The optimal α is:

$$\begin{aligned} \alpha \quad / \quad \tilde{u}_\alpha \approx u \\ \text{with} \quad \tilde{u}_\alpha = \text{argmin}_{\tilde{u}} J_\alpha(\tilde{u}) \end{aligned} \tag{34}$$

it was shown in [33] that the solution can be calculated with a fixed point algorithm, using a linearization of the gradient of the TV $\partial TV(u)$. At each step we linearize this term using the previous estimate $\tilde{u}_{n-1}$: $(\partial TV)_n \approx B_n = -\operatorname{div} \frac{\nabla}{|\nabla \tilde{u}_{n-1}|}$. Let B be this linear operator when the algorithm has converged. We have:

$$\tilde{u}_\alpha = (A^H A + \alpha\lambda B)^{-1} A^H w \tag{35}$$

by replacing in 34 :

$$\begin{aligned} u &\approx (A^H A + \alpha\lambda B)^{-1} A^H w \\ (A^H A + \alpha\lambda B)u &\approx A^H (Au + n) \\ \alpha\lambda ||Bu|| &\approx ||A^H n|| \\ \alpha\lambda &= \frac{||A^H n||}{||Bu||} \end{aligned} \tag{36}$$

We write the relationship between the extremal singular values of AA^H (which are the same as $A^H A$): $\sigma_{\min} = \frac{\sigma_{\max}}{\kappa}$ and suppose that B is independent from α. We find that $\alpha\lambda = \sqrt{a\frac{1}{\kappa} + b}$ were a, b are constants depending on A, B, u and n. We add the hypothesis that the optimal α is 0 with a conditioning of 1. This is equivalent as saying that no regularization is needed when no inversion is needed. It leads to $b = -a$ we identify the constant depending on the data $\sqrt{b}$ with λ ($\lambda = \sqrt{b}$). Consequently, $\alpha = \sqrt{1 - \frac{1}{\kappa}}$.
We choose λ which is minimizing the reconstruction noise.

C. Experiments

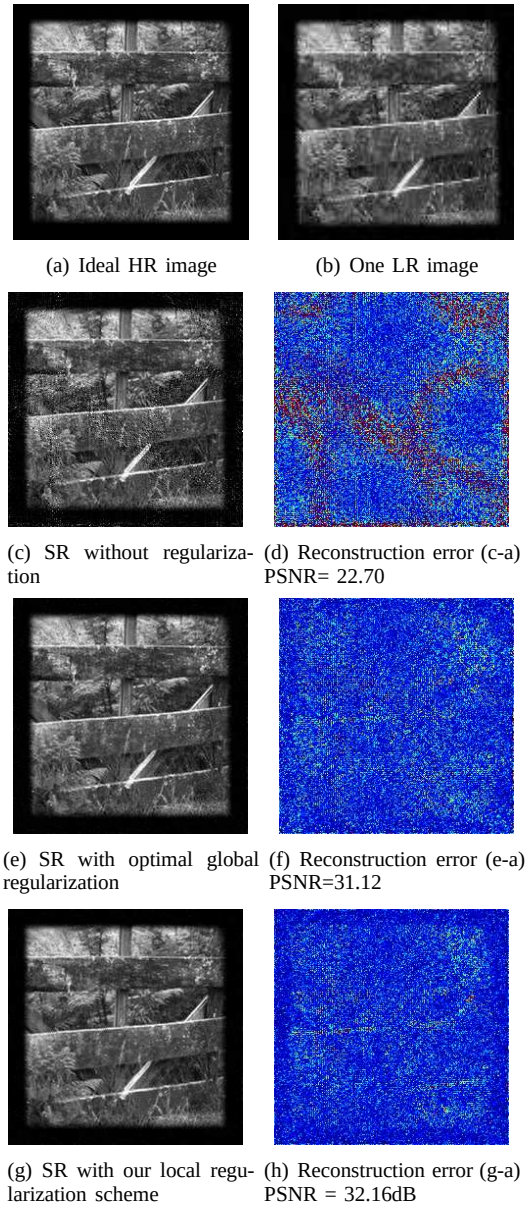
We show in Figure 5 how we predict local conditioning. We generate 4 noisy LR images from a 240×240 HR image (SR with $M = 2$, rotations between -5 and 5 degrees, translations distributed in $[0, M]^2$, zoom between $0.95, 1.05$) and perform SR interpolation without regularization, with optimal global TV regularization and optimal local regularization. global regularization case, the resulting image is excessively smoothed in areas with better conditioning and not enough elsewhere. With local regularization, the smoothing only occurs in the badly conditioned areas. This results in a better reconstruction of the HR image (in this example, 1dB PSNR gain).

D. Application to demosaicking

We show here that we can apply this algorithm for multi-image demosaicking. [34] showed the benefits of multi-image demosaicking. If the user has the opportunity to take 4 pictures of the same scene, we can use our optimal regularization scheme to reconstruct a HR image with $M = 2$ from the raw RGB components independently. We show in Figure 7 the comparison between a multi-image demosaicking with noise and a mono-image state of the art demosaicking (self-similarity driven demosaicking [35]) with noise. We generated a synthetic example by generating 4 LR versions a HR image and adding noise. The input for the mono-image demosaicking is a Bayer pattern generated from the HR image. Multi-image demosaicking with our local TV regularization gives a HR image without chromatic anomalies when compared to self-similarity driven demosaicking.

VI. Conclusion

We have studied super-resolution under a particular aspect. To avoid regularization (and subsequently hypotheses on images), we outlined contexts where we can perform unregularized SR. We began by giving an invertibility condition on the affine motion super resolution interpolation problem. We showed that little to no regularization is needed in the



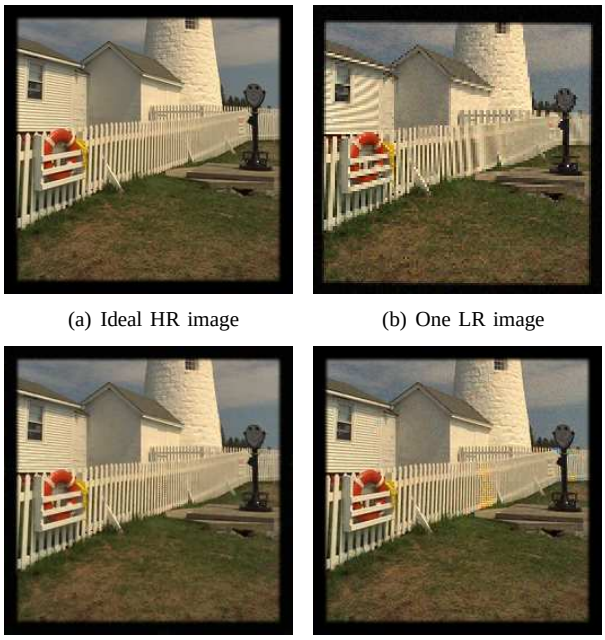
(a) Ideal HR image (b) One LR image

(c) SR without regularization (d) Reconstruction error (c-a) PSNR= 22.70

(e) SR with optimal global regularization (f) Reconstruction error (e-a) PSNR=31.12

(g) SR with our local regularization scheme (h) Reconstruction error (g-a) PSNR = 32.16dB

Fig. 6. *Local TV regularization for critical super resolution.* Reconstruction errors are shown in a blue-red color scale representing the gray level interval [0,30] (images are in [0,255]).



(a) Ideal HR image (b) One LR image

(c) Proposed multi-image demosaicking (d) Self-similarity driven demosaicking [35]

Fig. 7. *Multi-image demosaicking with 4 LR images using local regularization.* LR images are simulated from the ideal HR color image by sub-sampling each RGB component by a factor $M = 2$ after a random motion and adding noise with standard deviation $\sigma = 5$ (Pixels have values in $[0, 255]$).

context of affine motion SR with a large number of images, and more precisely that it is always in the best interest of SR interpolation to acquire more images. We also studied a critical case where regularization is necessary, but not everywhere. To minimize zones where holes are filled using the total variation term in the objective functional, we proposed a local conditioning measure which we used as local weights.

REFERENCES

[1] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *Int. J. Imaging Syst. Technol.*, vol. 14, no. 2, pp. 47–57, 2004.

[2] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, Sep. 2011.

[3] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using asequence of undersampled images," *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.

[4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[5] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient Fourier-Wavelet Super-Resolution," *Image Processing, IEEE Transactions on*, vol. 19, no. 10, pp. 2669–2681, Oct. 2010.

[6] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2830–2841, Nov. 2007.

[7] K.-H. Yap, Y. He, Y. Tian, and L.-P. Chau, "A Nonlinear -Norm Approach for Joint Image Registration and Super-Resolution," *Signal Processing Letters, IEEE*, vol. 16, no. 11, pp. 981–984, Nov. 2009.

[8] L. Alvarez, Y. Gousseau, and J.-M. Morel, *The Size of Objects in Natural and Artificial Images*. Elsevier, 1999, vol. 111, pp. 167–242.

[9] F. Malgouyres and F. Guichard, "Edge Direction Preserving Image Zooming: A Mathematical and Numerical Analysis," *SIAM Journal on Numerical Analysis*, vol. 39, no. 1, 2002.

[10] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *Image Processing, IEEE Transactions on*, vol. 18, no. 1, pp. 36 –51, jan. 2009.

[11] G. Peyré, S. Bougleux, and L. Cohen, "Non-local Regularization of Inverse Problems Computer Vision - ECCV 2008," in *Computer Vision - ECCV 2008*, ser. Lecture Notes in Computer Science, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2008, vol. 5304, ch. 5, pp. 57–68.

[12] Y. Zhuo, J. Liu, J. Ren, and Z. Guo, "Nonlocal based super resolution with rotation invariance and search window relocation," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, Mar. 2012, pp. 853–856.

[13] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Computer Vision, 2009 IEEE 12th International Conference on*, Oct. 2009, pp. 349–356.

[14] G. Rochefort, F. Champagnat, G. Le Besnerais, and J. F. Giovannelli, "An Improved Observation Model for Super-Resolution Under Affine Motion," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3325–3337, Nov. 2006.

[15] X. Zhang, J. Jiang, and S. Peng, "Commutability of Blur and Affine Warping in Super-Resolution With Application to Joint Estimation of

Triple-Coupled Variables," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1796–1808, Apr. 2012.
[16] A. Sánchez-Beato, "Coordinate-descent super-resolution and registration for parametric global motion models," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 1060–1067, Oct. 2012.
[17] K. Gröchenig and T. Strohmer, "Numerical and theoretical aspects of nonuniform sampling of band-limited images," in *Nonuniform sampling : Theory and Practice*, ser. Inf. Technol. Transm. Process. Storage. Kluwer/Plenum, New York, 2001, pp. 283–324.
[18] G. Facciolo, A. Almansa, J.-F. Aujol, and V. Caselles, "Irregular to Regular Sampling, Denoising, and Deconvolution," *Multiscale Modeling & Simulation*, vol. 7, no. 4, pp. 1574+, 2009.
[19] M. Arigovindan, M. Suhling, P. Hunziker, and M. Unser, "Variational image reconstruction from arbitrarily spaced samples: a fast multiresolution spline solution," *Image Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 450–460, 2005.
[20] A. Almansa, J. Caron, and S. Durand, "Deblurring of irregularly sampled images by TV regularization in a spline space," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, Sep. 2010, pp. 1181–1184.
[21] A. Papoulis, "Generalized sampling expansion," *Circuits and Systems, IEEE Transactions on*, vol. 24, no. 11, pp. 652–654, Nov. 1977.
[22] N. A. Ahuja and N. K. Bose, "Multidimensional Generalized Sampling Theorem for wavelet Based Image Superresolution," in *Image Processing, 2006 IEEE International Conference on*. IEEE, Oct. 2006, pp. 1589–1592.
[23] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
[24] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1185–1199, Sep. 2004.
[25] ——, "Statistical performance analysis of super-resolution," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1413–1428, Jun. 2006.
[26] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1+, Apr. 2003.
[27] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal Registration Of Aliased Images Using Variable Projection With Applications To Super-Resolution," *The Computer Journal*, vol. 52, no. 1, pp. 31–42, Jan. 2009.
[28] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
[29] F. Champagnat, G. Le Besnerais, and C. Kulcsár, "Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework," *J. Opt. Soc. Am. A*, vol. 26, no. 7, pp. 1730–1746, Jul. 2009.
[30] R. F. Bass and K. Grochenig, "Random sampling of multivariate trigonometric polynomials," *SIAM journal on mathematical analysis*, vol. 36, no. 3, pp. 773–795, 2005.
[31] H. Su, Y. Wu, and J. Zhou, "Super-Resolution Without Dense Flow," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1782–1795, Apr. 2012.
[32] Y. Traonmilin, S. Ladjal, and A. Almansa, "On the amount of regularization for Super-Resolution interpolation," in *20th European Signal Processing Conference 2012 (EUSIPCO 2012)*, Bucharest, Romania, Aug. 2012.
[33] T. F. Chan and P. Mulet, "On the Convergence of the Lagged Diffusivity Fixed Point Method in Total Variation Image Restoration," *SIAM J. Numer. Anal.*, vol. 36, no. 2, pp. 354–367, 1999.
[34] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *Image Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 141–159, Jan. 2006.
[35] Antoni Buades, Bartomeu Coll, Jean-Michel Morel, Catalina Sbert, "Self-Similarity Driven Demosaicking," *Image Processing On Line*, 2011.

VII. Appendix

A. Intermediate results for the invertibility

Lemma VII.1. *For $1 \leq i \leq N$, let $u_i \in \mathbb{C}^{n\times n}$, $u_i(r,s) = x_i^r y_i^s$, we call u_i 2D Vandermonde vectors with seed $[x_i, y_i]$. If for all $1 \leq i < j \leq N, x_i \neq x_j, y_i \neq y_j$, $\dim(\mathrm{span}(u_i)_{i=1,N}) = min(N, n^2)$.*

Proof: We show that the u_i are linearly independent if $N \leq n^2$. Let us suppose $\sum \lambda_i u_i = 0$. Let $u_i(s) = X_i y_i^s$ with $X_i = (x_i^r)_r$. for all s, $\sum \lambda_i u_i(s) = \sum \lambda_i X_i y_i^s = 0$. The X_i form an independent family of 1D Vandermonde vectors. It implies that $\sum \lambda_i y_i^s = 0$ which we rewrite $\sum \lambda_i Y_i = 0$, but the Y_i are also independent. Consequently, for all $i, \lambda_i = 0$. ∎

Lemma VII.2. *If for all $p_i, p_j \in \Gamma^c, ||q_1^{-1}p_i - q_2^{-1}p_j \mod 1||_0 = 2$, $\ker A_1 + \ker A_2 = \mathbb{C}^{ML\times ML}$.*

Proof: We prove this lemma for affine motions Q_i on finite discrete signals. Q_i^{-1} are performed by finite discrete Fourier interpolation. In practice, calculating Q_i the same way is a good approximation. We can construct a basis of $\ker A_1$ and $\ker A_2$ by taking the inverse transformations of the indicator functions of the pixels zeroed by the sub-sampling. In the Fourier domain, these bases are:

$$\hat{u}_i(\omega) = e^{-j\left\langle \omega,\, q_1^{-1}p_i\right\rangle},\ \hat{u}'_i(\omega) = e^{-j\left\langle \omega,\, q_2^{-1}p_i\right\rangle} \tag{37}$$

which are 2D Vandermonde vectors with seed $[e^{-j\left\langle e_x,\, q_k^{-1}p_i\right\rangle}, e^{-j\left\langle e_y,\, q_k^{-1}p_i\right\rangle}]$. We use Lemma VII.1: $\ker A_1 + \ker A_2 = \mathrm{span}((\hat{u}_i),(\hat{u}'_i)) = \mathbb{C}^{ML\times ML}$ (the seeds are all different because fo rall $p_i, p_j, ||q_1^{-1}p_i - q_2 p_j^{-1} \mod 1||_0 = 2$). ∎

Lemma VII.3. *Let $n < M^2$.If for all $p_i, p_j \in \Gamma^c, 1 \leq k_1 < k_2 \leq N, ||q_{k_1}^{-1}p_i - q_{k_2}^{-1}p_j \mod 1||_0 = 2$ and $\dim(\cap_{k=1,n} \ker A_k) = (M^2 - n)L^2$ then $\cap_{k=1,n} \ker A_k + \ker A_{n+1} = \mathbb{C}^{ML\times ML}$.*

Proof: Let (e_i) be a basis of $\cap_{k=1,n} \ker A_k$ of size $(M^2 - n)L^2$. In the basis $(u_j)_{j=1,n}$ of $\ker A_1$,:

$$e_i = \sum \alpha_{i,j} u_j \tag{38}$$

Let u'_i a basis of $\ker A_n$. With the hypothesis, any linear combination of e_i, u'_i is a linear combination of independent 2D Vandermonde vectors. Therefore, $\dim(\mathrm{span}((\hat{e}_i),(\hat{u}'_i))) = \min\left((ML)^2, (M^2-n)L^2 + (M^2-1)L^2\right) = (ML)^2$. Thus, we have $\cap_{k=1,n} \ker A_k + \ker A_2 = \mathrm{span}((\hat{e}_i),(\hat{u}'_i)) = \mathbb{C}^{ML\times ML}$. ∎

B. Intermediate calculation for the Hessian

Using conventional differentiation rules, we have at $\boldsymbol{\theta}_0$:

$$\begin{aligned}
\frac{\partial}{\partial\theta_i}[AA^\dagger] =& \frac{\partial}{\partial\theta_i}[A]A^\dagger + A\frac{\partial}{\partial\theta_i}[A^\dagger] \\
=& \frac{\partial}{\partial\theta_i}[A]A^\dagger - A(A^HA)^{-1}\frac{\partial}{\partial\theta_i}[A^HA]A^\dagger \\
&+ A(A^HA)^{-1}\frac{\partial}{\partial\theta_i}[A^H] \\
=& \frac{\partial}{\partial\theta_i}[A]A^\dagger - AA^\dagger\frac{\partial}{\partial\theta_i}[A]A^\dagger \\
&- (A^\dagger)^H\frac{\partial}{\partial\theta_i}[A^H]AA^\dagger + (A^\dagger)^H\frac{\partial}{\partial\theta_i}[A^H] \\
=& (I - AA^\dagger)\frac{\partial}{\partial\theta_i}[A]A^\dagger + (A^\dagger)^H\frac{\partial}{\partial\theta_i}[A^H](I - AA^\dagger) \\
=& 2Re((I - AA^\dagger)\frac{\partial}{\partial\theta_i}[A]A^\dagger)
\end{aligned} \tag{39}$$