FAST AND ACCURATE MULTI-FRAME SUPER-RESOLUTION OF SATELLITE IMAGES

Jérémy Anger[1,*], Thibaud Ehret[1], Carlo de Franchis[1,2], Gabriele Facciolo[1]

[1] Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 94235, Cachan, France
[2] Kayrros SAS

KEY WORDS: super-resolution, multi-frame, satellite images, spline, registration

ABSTRACT:

Recent constellations of small satellites, such as Planet's SkySats, offer new acquisition modes where very short videos or bursts of images are acquired instead of a single still image. Compared to sequences of multi-date images, these sequences of consecutive video frames yield a large redundancy of information within the range of seconds. This redundancy enables to increase the spatial resolution using multi-frame super-resolution algorithms. In this paper, we propose a novel super-resolution method based on a high-order spline interpolation model that combines multiple low-resolution frames to produce a high-resolution image. Moreover this method can be implemented efficiently on GPU to process entire images from real satellite acquisitions. Synthetic and real experiments show that the proposed method is able to recover fine details, and measurements of the resulting resolution indicate a gain of 10 cm / pixel with respect to Planet's SkySat standard imagery products.

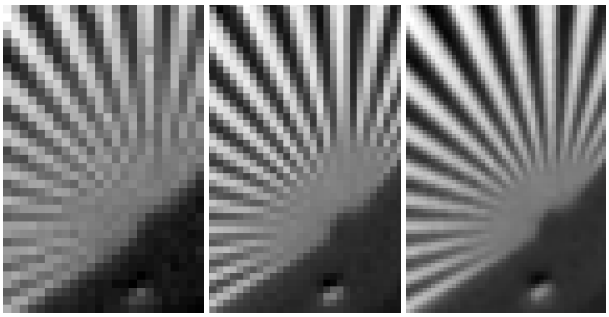
1. INTRODUCTION

Satellites play a big role in the observation of the Earth: from environmental monitoring to meteorology and even industry monitoring. Earth monitoring applications require a good ground resolution. For example, fine detection and analysis of human activity requires a resolution in the range of 30 cm to 1 m / pixel (Murthy et al., 2014).

Earth observation missions were historically owned by national organisations, constructing high-cost long-term satellites. Starting in the late 90's, a similar model was also adopted by actors from the private sector (*e.g.* IKONOS, EROS, QuickBird, WorldView). But in recent years, some companies have started to offer low-cost imagery thanks to new satellite designs. The current trend is to launch many smaller satellites with a shorter lifespan, providing a wider coverage at lower cost. For example, Planet provides a daily revisit time on some products. However, low-cost satellite means that quality of each individual image is lower, with higher noise or worse GSD for example. This means that instead of trying to obtain a 50 cm GSD from the physical design of the satellite, such resolution has to be reached using computational photography techniques and in particular multi-frame super-resolution.

This small satellite trend can be compared to the case of smartphone cameras: smartphones have lower grade optics and sensor compared to DSLR, but computational photography techniques improve the images to a satisfying quality. Indeed, smartphone manufacturers have been pushing the limits of the sensors by acquiring bursts of images and fusing them for joint denoising, demosaicing, and super-resolution.

The acquisition model for satellite images differs from regular cameras. Since the satellite is far enough from the ground, we can consider that the observed scene u lies on a plane at infinity. Then the projective camera model can be assumed to be an affinity denoted by A. The whole image formation process for the image v_i can therefore be summarized in a single equation:



Low-res. SkySat L1A frame | SkySat L1B | Proposed

Figure 1. Examples of reconstruction on a real ground fan target from 35 low-resolution SkySat L1A frames. From left to right: reference low-resolution L1A frame, L1B product provided by Planet ($\times 1.25$), proposed method ($\times 2$).

$$v_i = \Sigma_1((u \circ A_i) * k) + n_i, \tag{1}$$

where k is the Point Spread Function (PSF) modeling jointly optical blur and pixel integration, the operator Σ_1 is the bi-dimensional ideal sampling operator due to the sensor array, and n_i models the image noise. Since the satellite moves during the acquisition, small exposure times are necessary to avoid blur, thus the images are noisy. The resulting digital image v is encoded in a linear intensity scale, as is frequent in remote sensing applications. In the specific case of affine transforms, assuming that k is an isotropic kernel, A and k commute. Hence, it is possible to first estimate $u * k$, then invert k on the high-resolution image; this property allows for an efficient super-resolution method.

Note that in absence of noise, if the kernel k introduces a cutoff below the critical sampling frequency, then the image v is said to be well sampled. Therefore, by the Nyquist–Shannon sampling theorem (Shannon, 1948) the continuous signal $u * k$ could be reconstructed from the samples in v. In this scenario, there would be no gain in applying multi-frame super-resolution on this system, except for denoising. The interesting case is

*Corresponding author (angerj.dev@gmail.com)

when the system is designed to have the frequency cutoff of k above the critical sampling frequency. The resulting image v is then said to be aliased. Aliasing is the phenomenon of replication of higher frequencies onto lower frequencies. Supermode (Latry, Rougé, 2000) for SPOT5 used this information to produce a higher resolution image from two images acquired simultaneously with a half-pixel shift. The objective of multi-frame super-resolution is to collect samples from multiple different images to estimate a de-aliased $u * k$.

The design of the SkySat-1 satellite (Murthy et al., 2014) from Planet participates in the trend towards small but high-resolution satellites. Super-resolution played an important role in the design of the satellite and influenced the optics design. Indeed, the low-level images are not only aliased, they are also acquired in bursts. Compared to the traditional push-broom cameras, SkySat-1 contains a full-frame sensor and is able to capture high-definition videos. This means that higher quality images can be computed directly by combining frames from a single SkySat acquisition. Figure 1 shows an individual frame (L1A) and the provided super-resolved product (L1B).

In this paper we present a novel super-resolution method based on a robust registration followed by a non-uniform sampling fusion step. We use an inverse compositional technique for the registration and a fast GPU implementation of the ACT algorithm (Feichtinger et al., 1995) adapted to use a spline image model for the fusion. We also show the improvements on SkySat images over the commercial product currently proposed by Planet.

In the rest of the paper, we first review related works on multi-image super-resolution, in particular for remote sensing. We then describe the method, comprised mainly of a registration and a fusion step. Finally, experiments are carried out on synthetic and real images, highlighting the performance of our method with respect to existing super-resolution methods.

2. RELATED WORKS

Super-resolution is an important problem of image processing. There have already been many reviews presenting the problem in detail such as the recent (Nasrollahi, Moeslund, 2014, Yue et al., 2016). These reviews classify methods into two major categories: the methods that do super-resolution using only the reference image (single image super-resolution) and multi-image super-resolution (for example from bursts or videos). Here, we only focus on the case of multi-image super-resolution. Indeed these reviews have shown that having multiple input images actually increases the quality of the reconstruction. It is also important to note that pan-sharpening (Garzelli, 2016) is not super-resolution. Indeed, for pan-sharpening a higher-resolution guide is available, which is not the case in our application.

Four classic categories of multi-image super-resolution methods can be identified: kernel regression, shift-and-add, variational, and spectral methods.

In kernel regression methods (Takeda et al., 2007, Takeda et al., 2009), the pixels of the high-resolution image are computed by solving a weighted linear least squares problem. The contribution of the samples (expressed in a common coordinate system after registration) are limited to a small spatial neighborhood and are adjusted by weights derived from a kernel. Usually the weights only take into account the spatial distance of the samples to the estimated pixel. However, more recent methods try also to take into account the radiometric information or the local structure (Wronski et al., 2019).

Shift-and-add methods produce a high resolution image by registering several low resolution images and integrating the pixels of the low resolution frames onto the high resolution one. For some methods, a low resolution pixel affects only one high resolution pixel at its nearest neighboring location (Keren et al., 1988, Farsiu et al., 2004, Murthy et al., 2014), while for others it affects the area of high resolution pixels covered by the low resolution pixel (Merino, Nunez, 2007). The samples are usually averaged (using weights or not), or obtained by robust aggregation (such as the median) to remove outliers (Farsiu et al., 2004, Murthy et al., 2014). Earlier methods assumed that enough images were aggregated so that the result had no holes (Keren et al., 1988). To fill-in holes and also to remove outliers, regularizers based on the Total Variation (TV) (Farsiu et al., 2004) are frequently used within an energy minimization post-process. Once the high resolution image is produced, the methods usually have a last step to remove the blur introduced by the PSF (Murthy et al., 2014). Shift-and-add methods can be seen as simple variational methods.

Variational methods (Tom, Katsaggelos, 1995, Marquina, Osher, 2008) solve the super-resolution problem by minimizing a cost function. The low resolution images are usually first registered. However, the registration can also be refined during the minimization process (Tom, Katsaggelos, 1995, Peng et al., 2012). The cost function integrates a data term and a prior term. The data term is typically the distance between the samples and simulation of the low-resolution images from the current estimated high-resolution image using the camera model (blur, downsampling, motion). A frequent prior used for these methods is TV (Marquina, Osher, 2008). In addition to having registration refined during the minimization process (such as alternating one step of minimization and one step of refinement of the registration), other parameters like the blur kernel can be refined the same way. For variational methods, the cost function is usually justified by a Bayesian model (Tom, Katsaggelos, 1995). Projection onto convex sets (POCS) (Tekalp et al., 1992) can be considered as a variational method.

Contrary to the previous methods, spectral methods combine the images in a transformed domain, such as Fourier. Kim et al. (Kim et al., 1990) proposed a recursive method based on a weighted least squares optimization problem expressed in the Fourier domain. While Fourier is the most used, other bases have also been used with success. For example, Nguyen and Milanfar (Nguyen, Milanfar, 2000) proposed an iterative reconstruction in the wavelet domain to minimize a Tikhonov-regularized least squares problem. A limitation of such methods is that the reconstruction task can be significantly ill-conditioned when the sampling is degenerate. This can impact the entire image depending on the basis used (such as Fourier) even if the problem is only very local.

More recently, many methods based on neural networks have been proposed for natural image and video super-resolution. These methods, such as (Lim et al., 2017, Dai et al., 2019, Shi et al., 2016, Wang et al., 2019), have been shown to work very well on natural images. Recently the European Space Agency also organized a challenge (Märtens et al., 2019) with the objective of super-resolving images coming from the PROBA-V satellites. For this contest, the winning method (Molini et al.,

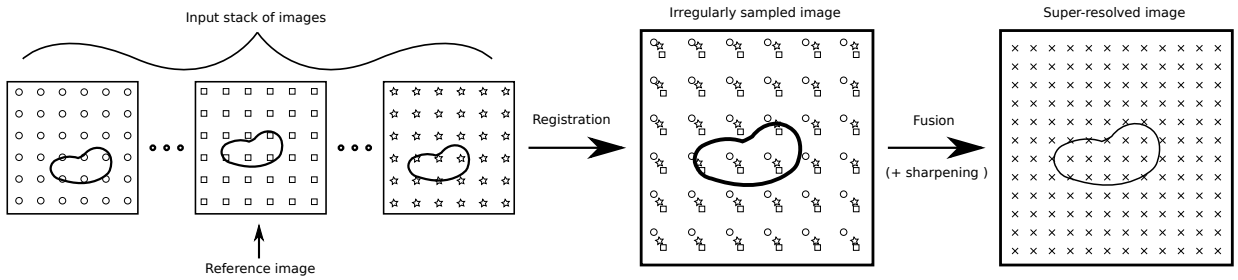


Figure 2. Pipeline of the super-resolution method. Each input image is registered onto a reference image to create an irregularly sampled signal. The uniformly sampled image at the requested resolution is then produced by combining the samples.

2019) is a neural network. The specificity of this challenge is the availability of ground truth images. Indeed, there are two different generations of the PROBA-V satellite with different sensor resolutions, therefore it is possible to train networks to produce images similar to high-resolution acquisitions from the low-resolution ones. For most satellites, such ground truth data is not available.

In general, however, it is difficult to apply neural networks for remote sensing super-resolution. As it was shown in (Wang et al., 2019), having the correct training data for neural networks is very important; otherwise, the networks suffer from dataset bias. This means that the usual training datasets cannot be used for remote sensing. Simulating data is also very difficult as the down-sampling kernel is usually unknown. For example, the problem usually considered with neural networks is a down-sampling by a factor of four with a bicubic kernel which is unrealistic because it neglects the optical system and sensor integration. Moreover, since no information is available to restore such high frequencies, the evaluation criterion is often the perceptual quality. However, it is very important to not introduce any bias into the estimation (such bias is usually called "hallucinations"). In particular, hallucinating can be very problematic in critical applications such as defense. This is why neural networks are usually avoided for restoring remote sensing data.

3. PROPOSED METHOD

As most multi-frame super-resolution algorithms, our method is based on two main steps: registration and fusion. The registration step estimates a subpixel accurate affine deformation between the low-resolution frames. Using this precise information, our fusion step produces a high-resolution image using a 2D spline model. These two steps are illustrated in Figure 2 and are described in detail in the following sections. Additionally, depending on the characteristics of the optical system, a sharpening step can be applied as post-processing.

3.1 Registration

The first step of the pipeline estimates an affine transformation between each frame and a given reference frame. While the popular option to align two images usually relies on dense optical flow, it is not the most adapted for aligning satellite images. Indeed, as all elements are considered to be in the same plane at infinity, the movement between two satellite images can be assumed to be an affinity. Using such parametrization allows for a more robust and precise alignment than using an optical flow since only six coefficients per image are estimated instead of two per pixel for a dense optical flow.

The affinity between two images is estimated using the inverse compositional algorithm (Baker, Matthews, 2001), which has been shown to be very precise as well as being robust to noise (Sánchez, 2016). More details about the method can be found in (Briand et al., 2018). As a minor modification, we replaced the bicubic interpolation by a spline interpolator of order 3, which slightly reduces the computational cost.

Since the images are aliased, it is important to first apply a low-pass filter (a simple Gaussian blur in our case) before registration. It has been shown (Vandewalle, 2006) that aliasing could lead to misregistration, causing a loss of resolution down the pipeline.

Due to the motion of the satellite during the acquisition of a burst, there may remain little overlap between the first (or last) image of the sequence and the reference one, and we found that the estimation of a global transform in these cases is prone to errors. Instead, by computing the expected overlap using a prior of the satellite motion, we detect frames with less than 80% overlap with the reference frame and use intermediate frames to register them by composing the estimated transforms. While this composition could result in error accumulation, we found that estimating the displacement between images with little overlap was prone to larger errors and instabilities.

The result of the registration process is one affinity per input frame which gives a sub-pixel position to each sample relative to the high-resolution coordinate system, as illustrated in Figure 2. In this common coordinate system, the pixels from the low-resolution frames represent an irregular sampling. The next section describes how to combine these samples to create a high-resolution image.

One limitation of using a global parametrization for the registration is related to the parallax effect. As tall structures break the assumption that the observed objects are in the same plane, they appear with a slight offset. In practice, given the speed and altitude of the satellite the impact of parallax is very small (below the pixel size). Indeed, with a satellite altitude of $H \simeq 500$ km and a baseline (distance between two camera centers) between consecutive frames of $B \simeq 150$ m, the parallax for a structure of altitude $z = 25$ m is approximately $\frac{B}{H} z \simeq 0,00725$ m $\simeq 0,00725$ pixels per frame (assuming approximately a 1 m/pixel resolution). As can be seen in Figure 3, the effect of this amount of parallax error is negligible (comparable to a small amount of noise). Nevertheless, in case of abrupt elevation changes in the imaged surface (tall building or cliff), corrective offsets should be estimated, for example by using a local multi-frame motion estimation approach (Rais et al., 2016).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-1-2020-57-2020 | © Authors 2020. CC BY 4.0 License.

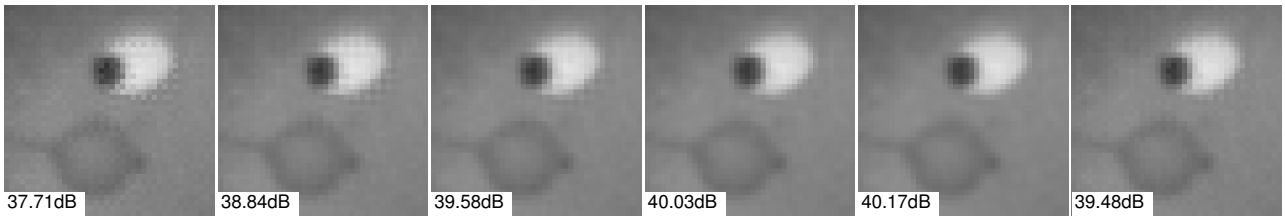


Figure 3. Crops from the fusion of 15 frames with Gaussian noise of standard deviation $\sigma = 2/255$ and ground-truth registration except for one frame with an offset of $(+1, +1)$ pixel (zoom $\times 2$). Notice how the artifacts due to the misregistration disappear with high-orders spline interpolation.

3.2 Fusion with ACT using trigonometric polynomial interpolation

The goal of a fusion method for multi-frame super-resolution is to combine the samples and their estimated positions into a high-resolution image. The proposed fusion method is based on the ACT algorithm (Strohmer, 1995) and spline interpolation. In order to properly describe it, we first review in this section the ACT algorithm with trigonometric polynomial interpolation as originally proposed by Gröchenig and Strohmer. We explain our adaptation of ACT for the spline model in the following section.

The Adaptive weights Conjugate gradient Toeplitz (ACT) method (Strohmer, 1995, Strohmer, 1997) was designed to reconstruct an image from irregular samples. In the case of super-resolution, we consider the aligned low-resolution frames as a set of irregular samples and aim to resample it into a regularly sampled image. The ACT method models a discrete image u as a trigonometric polynomial of order $\frac{M}{2} \times \frac{M}{2}$ (with M even and positive), so that the interpolation at the irregular sampling positions $\Xi = \{\boldsymbol{\xi}_k\}_{k=1}^K \subseteq \mathbb{R}^2$ becomes

$$u(\boldsymbol{\xi}_k) = \sum_{\boldsymbol{t} \in \{-\frac{M}{2}+1, \dots, \frac{M}{2}\}^2} \hat{u}_{\boldsymbol{t}} e^{\frac{2\pi i}{M} \cdot \langle \boldsymbol{\xi}_k, \boldsymbol{t} \rangle}, \; k \in \{1, \dots, K\}, \tag{2}$$

where $\{\hat{u}_t\}_{t=1}^{M^2}$ are the coefficients of the trigonometric polynomials. Thus, denoting z the irregularly sampled data at positions Ξ, the forward model is written as

$$z = S\hat{u}, \;\; \text{with } S = ((s_{k\boldsymbol{t}})), \; s_{k\boldsymbol{t}} = e^{\frac{2\pi i}{M} \langle \boldsymbol{\xi}_k, \boldsymbol{t} \rangle}. \tag{3}$$

The operator S evaluates the trigonometric polynomial at positions Ξ and can be applied using the nonequispaced Fast Fourier Transform (nFFT) (Potts et al., 2001).

The objective of ACT is to recover the coefficients $\hat{u}$ from the samples z by solving the following least squares problem

$$\arg\min_{\hat{u}} \|\sqrt{W}(S\hat{u} - z)\|_2^2, \tag{4}$$

where the optional diagonal matrix W acts as a pre-conditioner that assigns weights to the samples that are inversely proportional to the local sampling density (Feichtinger et al., 1995, Facciolo et al., 2009). In practice this density per sample is estimated from the area of the Voronoi cell associated with said sample. The normal equation corresponding to Equation (4), omitting the weights, is

$$S^* S\hat{u} = S^* z. \tag{5}$$

This linear problem is solved with Conjugate Gradient and regularized by early stopping its iterations. The term S^*z is computed using the nFFT and the application of S^*S can be accelerated by observing that it is a Toeplitz matrix which can be made circulant and thus diagonal in Fourier (Feichtinger et al., 1995). Thanks to this property, each iteration of Conjugate Gradient only requires the application of two Fast Fourier Transforms. After the estimation of $\hat{u}$, the inverse discrete Fourier transform is applied to recover u at regularly-spaced positions.

It is important to notice that depending on the number of samples, the sampling pattern, and the desired zoom factor, the problem can be under-determined. One way to render the estimation well-posed is to reduce the bandwidth of the trigonometric polynomial. This can be achieved by manually choosing a lower degree polynomial for the reconstruction or by estimating a restricted spectral support as with the heuristic proposed in (Facciolo, 2011). Here, we will assume that the available samples are sufficient for solving the problem.

3.3 Fusion with ACT using spline interpolation

Trigonometric polynomials, as used in the method presented in Section 3.2, are global interpolators. Since our data can contain outliers with respect to the assumed model, this non-locality implies that errors will propagate spatially everywhere throughout the result. Moving objects, misregistrations and invalid boundary conditions are examples of such outliers. As a small experiment, the rightmost image in Figure 3 shows the reconstruction using trigonometric polynomials on a synthetic burst containing one outlier frame due to a simulated mis-registration. The reconstruction shows artifacts around edges, which are mitigated when using spline interpolation. Furthermore, due to the global nature of trigonometric polynomials, the computational cost of the method makes it unattractive for processing large images.

Instead, we propose to use a more local interpolator such as B-splines. For a given image u, its n-th order spline interpolation $\varphi^{(n)}$ is defined by

$$\varphi^{(n)}(x_1, x_2) = \sum_{i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}} c_{i_1,i_2} \beta^{(n)}(x_1 - i_1) \beta^{(n)}(x_2 - i_2), \tag{6}$$

with $c = (c_{i_1,i_2})$ defined such that $\varphi^{(n)}(k_1, k_2) = u(k_1, k_2)$, $\beta^{(i)}$ as

$$\beta^{(0)}(x) = \begin{cases} 1 & \text{when } -1/2 < x < 1/2 \\ 1/2 & \text{when } x = \pm 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and

$$\beta^{(n+1)} = \beta^{(n)} * \beta^{(0)}. \tag{8}$$

An in-depth study of B-spline interpolation can be found in (Briand, Monasse, 2018).

Since $\varphi^{(n)}$ is continuous, it can be used to define u at an irregular sampling position $\boldsymbol{\xi}$ with

$$u(\boldsymbol{\xi}) = \varphi^{(n)}(\boldsymbol{\xi}). \tag{9}$$

Similarly to Equation 3, the irregularly sampled data z at position $\Xi = (\boldsymbol{\xi}_{1,k}, \boldsymbol{\xi}_{2,k})$ can be written

$$z = Bc, \tag{10}$$

where c is the vector of spline coefficients associated to u and $B = (\beta^{(n)}(\boldsymbol{\xi}_{1,k} - i_1)\beta^{(n)}(\boldsymbol{\xi}_{2,k} - i_2))_{k,(i_1,i_2)}$. In practice, both B and c are finite due to computation precision.

We model the spline-based fusion with the ACT equation by replacing the Fourier coefficients with spline coefficients. The energy associated with Equation (10) is

$$\arg\min_c \|z - Bc\|_2^2. \tag{11}$$

This optimization problem can be solved using Conjugate Gradient, with the corresponding normal equation

$$B^T Bc = B^T z, \tag{12}$$

where B^T denotes the transpose of B. After c has been estimated, the spline coefficients are used to sample the image u by interpolating values on the desired regular grid.

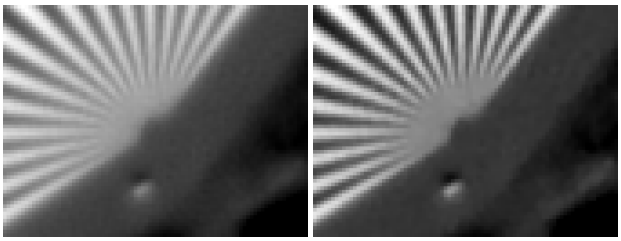
Unlike the Fourier case, $B^T B$ has no remarkable structure, therefore the operators B^T and B are successively applied. However, since B is sparse its application is much more efficient than its Fourier counterpart. Since these operations are local, they can be implemented efficiently in GPU. Using a fast parallel spline implementation (Briand, Davy, 2019), we are able to fuse large images within few seconds, compared to minutes with the trigonometric polynomial model. We found that 20 iterations of Conjugate Gradient is a good compromise between reconstruction quality and running time. Furthermore, Figure 3 shows the same reconstruction with different spline orders and we observe that high-order splines act as effective regularizers. Yet, as high-order splines are more costly to evaluate, we choose to use the order 9 for the rest of the paper.

3.4 Image sharpening

We recall that the objective of the fusion step is to combine samples to produce a high-resolution – but blurry – image $u * k$, where k includes the pixel integration as well as the optical transfer function. Let us call u_b the image that results from the fusion. The image sharpening aims to invert the blur introduced by k during the acquisition. However, as we do not know the optical characteristics, we empirically design a blur kernel k' such that the reconstruction is sharp, well-contrasted and with a low noise level.

We design the blur kernel k' defined as attenuation in the frequency domain comprised of two main components:

$$\mathcal{F}(k')(\boldsymbol{\omega}) = C(|\boldsymbol{\omega}|) \cdot S(\boldsymbol{\omega}). \tag{13}$$



Before sharpening After sharpening

Figure 4. Illustration of the effect of the sharpening step on the reconstruction. The image on the right (with sharpening) is more contrasted and less noisy than on the left (before sharpening).

The function C is the radial contrast attenuation due to optical blur, and we model it as $C(r) = (ar + 1)^{-1}$, with $a = 3.5$ in our experiments. The second part S relates to the sensor spatial integration during acquisition. In particular, S is modeled as the transition in the frequency domain between the integration filter of the pixels of size z and the filter corresponding to pixels at the finest grid (of size 1):

$$S(\boldsymbol{\omega}) = \operatorname{sinc}(z\boldsymbol{\omega}_x)\operatorname{sinc}(z\boldsymbol{\omega}_y) \cdot (\operatorname{sinc}(\boldsymbol{\omega}_x)\operatorname{sinc}(\boldsymbol{\omega}_y))^{-1}, \tag{14}$$

where $\operatorname{sinc}(\cdot)$ corresponds to the frequency response of the pixel integration.

In order to invert the blur kernel k', we formulate the following non-blind deconvolution problem and seek to restore u from u_b:

$$\arg\min_u \|u \cdot k' - u_b\|_2^2 + \lambda_1 \|\nabla u\|_1 + \lambda_2 \|\nabla u\|_2^2, \tag{15}$$

where λ_1 and λ_2 offer a balance between the TV regularization (denoises and favors sharpening) and the Tikhonov regularization on the gradients to avoid the staircasing effect of TV. We set $\lambda_1 = 0.4$ and $\lambda_2 = 10^{-2}$. This inverse problem can be solved efficiently using a half-quadratic splitting method (Krishnan, Fergus, 2009), and extended to noisy image deblurring using (Anger et al., 2019). Figure 4 illustrates the effect of sharpening on real data. The image after sharpening is both less noisy and more contrasted as a result of the inversion of the blur and the regularization.

4. EXPERIMENTS

In this section we first compare the proposed methods with different methods presented in Section 2 on synthetic data. Then, we compare our results obtained on real data acquired with the SkySat satellites with the super-resolved L1B product distributed by Planet. In the synthetic experiments the ground truth is available, which allows to use a standard metric such as PSNR to measure the quality of the reconstructions. To assess quantitatively the quality of the results on real images without ground truth we estimate the modulation transfer function (MTF) on a slanted-edge target. The information provided by the MTF combined with the estimated SNR allows to establish a notion of resolution, which is used to compare super-resolved results.

The slanted-edge method for MTF estimation was proposed by (Reichenbach et al., 1991), standardized as ISO 12233 (ISO, 2014) and later improved by (Roland, 2015). It collects samples along a strong edge, which requires to estimate its precise angle, and accumulates the samples into a finer grid. This results in a

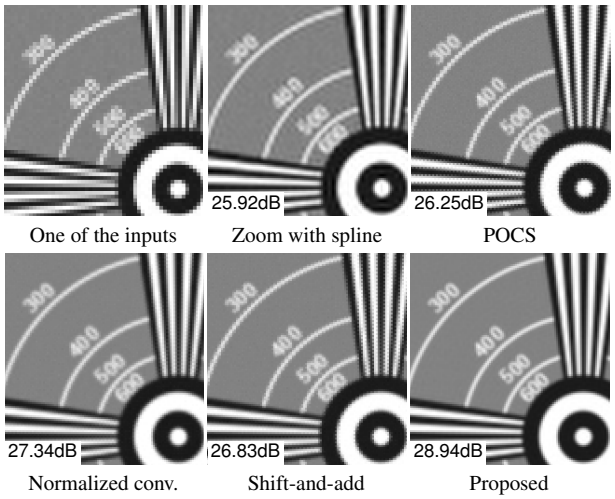


Figure 5. Synthetic example of super-resolution on the EIA Target (zoom $\times 2$). The figures correspond to the methods described in the text.

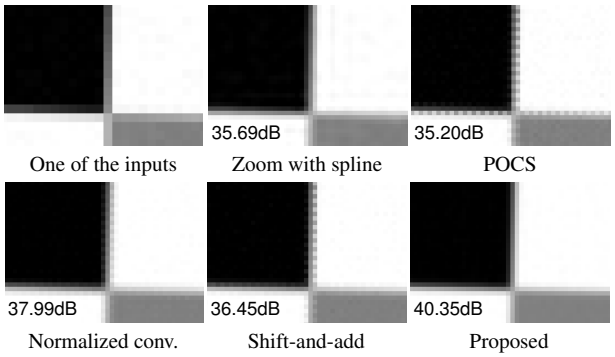


Figure 6. Super-resolution of a synthetic slanted-edge target for estimating the MTF (zoom $\times 2$). The MTFs estimated on this example are presented in Figure 7. The figures correspond to the methods described in the text.

denoised high-resolution 1D profile of the edge. The derivative of the edge yields the line-spread function, which in turn is used to compute the MTF by taking the modulus of its discrete Fourier Transform. This method allows to assess the per-frequency contrast reduction of the system assuming a radial symmetry of the MTF and without taking noise into account. In a noiseless world this attenuation could be undone for all the frequencies as done in Section 3.4. However, as we will see, in the real world noise will ultimately limit the resolution.

4.1 Super-resolution on synthetic data

In this section we compare the proposed method (ACT-spline) to other classic super-resolution methods. In particular we compare to a zoom using splines of order 9 with the implementation of (Briand, Davy, 2019), to shift-and-add (Keren et al., 1988), to POCS (Tekalp et al., 1992) and to normalized convolution (Takeda et al., 2007). Since these methods only assume a translation between the input images, all experiments in this section have been generated using only translations. Real data, that do not follow a translation model, are presented in Section 4.2. For both experiments, we first generated a well sampled image at the required resolution from a vector graphic. We then used this image to generate the burst of images with random translations, subsampled by a factor two with pixel integration and

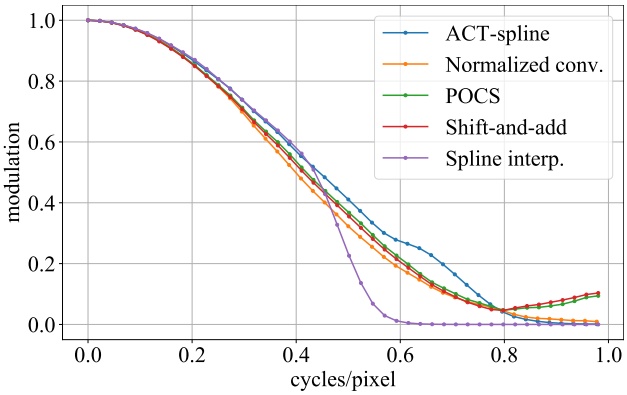


Figure 7. MTF estimated on the reconstruction of the synthetic slanted-edge target ($\times 2$) using different methods (shown in Figure 6). The proposed method (ACT-spline) has a higher MTF than the other methods, which means that the method was able to recover more information in high frequencies.

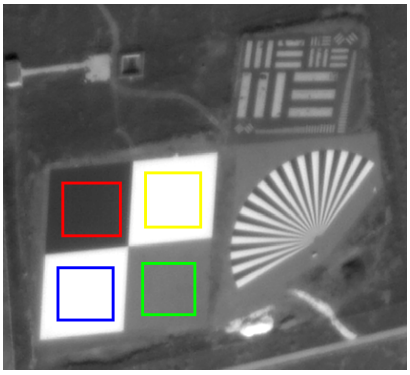


Figure 8. We evaluate the quality of the proposed method on real data using this target located in Baotou, China. The regions indicated by the four colors are used to estimate the SNR in the reconstructed images. The SNR is estimated independently for each region and is reported in Table 1.

added Gaussian noise of standard deviation $2/255$. The generated bursts consist of 18 images each. Moreover, we did not apply the sharpening step since we compare specifically the fusion methods without post-processing.

The first experiment was generated using a crop of the EIA resolution chart. Visual results as well as PSNRs are shown in Figure 5. Normalized convolution and ACT-spline are the only methods with no visual artifacts, however it is easier to read the small characters such as the 600 in the result from the proposed method. The same observation can be made by looking at the PSNR where ACT-spline is the best method followed by normalized convolution.

The second experiment was generated using a synthetic slanted-edge target. The goal of this experiment is to generate MTFs for the different methods. Results of the super-resolution are shown in Figure 6 and the corresponding MTFs are shown in Figure 7. We conclude that overall ACT-spline performs better than the other methods as its MTF is always above the other MTFs, especially in high-frequencies. The high frequency "bounce" observed in the MTFs of POCS and shift-and-add are caused by the artifacts in their reconstructions.

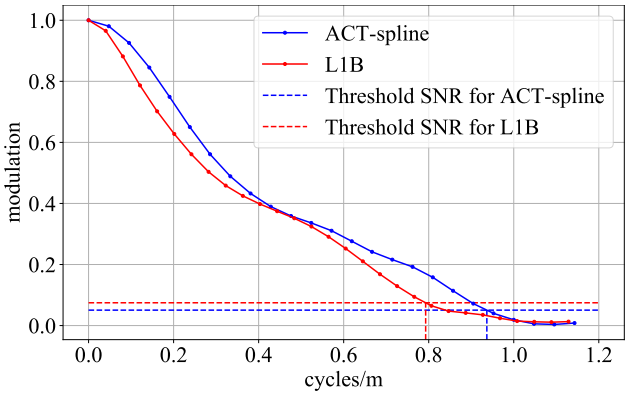


Figure 9. MTF estimated using a SkySat acquisition of a real calibration target located in Baotou, China. The proposed method ACT-spline produces an image with sharper edges than the L1B product provided by Planet. Using this plot a ground resolution per pixel can be estimated from the images: L1B is 63 cm/pixel while ACT-spline is 53 cm/pixel.

4.2 Super-resolution on real satellite images

Since ground truth is not available for real satellite images, the quantitative measure of the quality of the super-resolution is performed using a slanted-edge target. Thankfully, such are available, for example in the calibration site of Baotou in China (Li et al., 2015). The target, shown in Figure 8, can be used to evaluate both qualitatively, using the fan target and USAF resolution test patterns, and quantitatively, using the slanted-edge target.

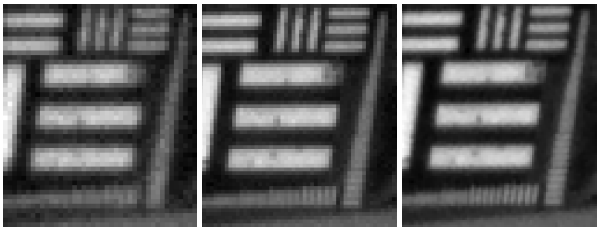
In this section, the super-resolved images are produced from SkySat low-resolution images (L1A), comprising of 35 frames with an average overlap of 17 samples per low-resolution pixel due to the motion of the satellite. We compare two methods: ACT-spline and the L1B produced by Planet which corresponds to a $\times 1.25$ zoom factor. In this section, we compare the two methods after sharpening since we want to compare the quality of the final product, *i.e.* at the end of the pipeline. We do not compare with the other methods from Section 4.1 since we have shown that our method performs better. Moreover most of these methods assume a translation between the input images which is not the case for real satellite images.

The results on the fan target are shown in Figure 1. Visually, the bands of the fan-shaped target are longer and sharper for the proposed method than the L1B image: ACT-spline has a better ground resolution. The same can be seen on the USAF resolution test patterns shown in Figure 10.

To verify quantitatively this observation, we used the slanted-edge target. From this target we first estimate the MTF using the same technique as for the synthetic target in Section 4.1. The MTF is shown in Figure 9. Our MTF is above the MTF of the L1B product which means that our reconstruction is sharper. This plot can also be used to estimate a ground resolution per pixel. For that we first estimate the SNR of the two reconstructed images on homogeneous areas, the results are presented in Table 1. We can see that the proposed result is less noisy. The ground resolution can be estimated by defining a reference energy and verify for how many cycles per meter the MTF crosses this energy. This can then be transformed to cm/pixel. For that we used the energy corresponding to two times the worst SNR reported in Table 1 for the two methods. This leads to a

		Signal level	Noise level	SNR
Zone 1 (red)	Low-res.	690.74	12.92	53.46
	L1B	690.64	10.00	69.02
	Proposed	688.82	6.45	**106.76**
Zone 2 (yellow)	Low-res.	1851.36	19.91	93.00
	L1B	1846.88	13.31	138.68
	Proposed	1887.48	13.26	**142.24**
Zone 3 (blue)	Low-res.	1855.39	18.57	99.88
	L1B	1846.91	12.62	**146.39**
	Proposed	1892.45	13.49	140.30
Zone 4 (green)	Low-res.	967.13	15.17	63.75
	L1B	963.63	12.37	77.89
	Proposed	971.39	10.67	**91.02**

Table 1. Signal-to-noise ratio (SNR) estimated on the SkySat images of Figure 8 before and after super-resolution for two methods (L1B provided by Planet and our ACT-spline). Super-resolution always improves the SNR and ACT-spline performs almost always the best for all signal levels.



Low-res. L1A frame L1B Proposed

Figure 10. Examples of reconstruction on real resolution test patterns from 35 low-resolution SkySat L1A frames. From left to right: Reference L1A frame, Planet L1B reconstruction ($\times 1.25$), proposed method ($\times 2$).

ground resolution of 63 cm/pixel (0.79 cycles/m) for the L1B product and 53 cm/pixel (0.94 cycles/m) for ACT-spline. This shows that our method produces images with a better quality than the L1B images provided by Planet. Moreover, the proposed method produces a $\times 2$ super-resolution image from 35 images (2560×1080) in less than a minute, including 30 s for the registration (CPU), 13 s for the fusion (GPU) and 5 s for the sharpening (CPU). As the registration is also GPU-friendly, it would be possible to significantly speed-up the method.

5. CONCLUSION

In this work we have proposed a novel super-resolution method for satellite images. The method is based on a noise robust registration step and a fusion based on a spline interpolation model. Moreover, the fusion step can be implemented on GPU making the method a viable practical alternative to current techniques. Using real images from the SkySat constellation, we have shown that this method can create images with a higher effective ground resolution than current commercial products. We have measured a gain of 10 cm / pixel with respect to Planet's SkySat L1B product, while improving the signal to noise ratio. The method can still be improved by taking into account parallax and by handling moving objects in the registration.

ACKNOWLEDGEMENTS

Work partly financed by the Office of Naval Research grant N00014-17-1-2552, CNES MISS project, DGA Astrid project ≪filmer la Terre≫ n° ANR-17-ASTR-0013-01, and Kayrros. We thank Planet for providing the L1A SkySat images.

REFERENCES

Anger, J., Delbracio, M., Facciolo, G., 2019. Efficient blind deblurring under high noise levels. *IEEE ISPA*, 123–128.

Baker, S., Matthews, I., 2001. Equivalence and efficiency of image alignment algorithms. *IEEE CVPR*, 1, I–1090.

Briand, T., Davy, A., 2019. Optimization of Image B-spline Interpolation for GPU Architectures. *IPOL*, 9, 183–204.

Briand, T., Facciolo, G., Sánchez, J., 2018. Improvements of the Inverse Compositional Algorithm for Parametric Motion Estimation. *IPOL*, 8, 435–464.

Briand, T., Monasse, P., 2018. Theory and Practice of Image B-Spline Interpolation. *IPOL*, 8, 99–141.

Dai, T., Cai, J., Zhang, Y., Xia, S.-T., Zhang, L., 2019. Second-order attention network for single image super-resolution. *CVPR 19*, 11065–11074.

Facciolo, G., 2011. *Irregularly sampled image restoration and interpolation*. Universitat Pompeu Fabra.

Facciolo, G., Almansa, A., Aujol, J.-F., Caselles, V., 2009. Irregular to Regular Sampling, Denoising, and Deconvolution. *Multiscale Modeling & Simulation*, 7(4), 1574–1608.

Farsiu, S., Robinson, D., Elad, M., Milanfar, P., 2004. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2), 47–57.

Feichtinger, H. G., Gr, K., Strohmer, T. et al., 1995. Efficient numerical methods in non-uniform sampling theory. *Numerische Mathematik*, 69(4), 423–440.

Garzelli, A., 2016. A review of image fusion algorithms based on the super-resolution paradigm. *Remote Sensing*, 8(10), 797.

ISO, 2014. Resolution and spatial frequency response. Standard, International Organization for Standardization.

Keren, D., Peleg, S., Brada, R., 1988. Image sequence enhancement using sub-pixel displacements. *CVPR 88*, 742–743.

Kim, S., Bose, N. K., Valenzuela, H. M., 1990. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE ICASSP*, 38(6), 1013–1027.

Krishnan, D., Fergus, R., 2009. Fast image deconvolution using hyper-laplacian priors. *NIPS*, 1033–1041.

Latry, C., Rougé, B., 2000. Optimized sampling for ccd instruments: the supermode scheme. *IGARSS*, 5, IEEE, 2322–2324.

Li, C., Tang, L., Ma, L., Zhou, Y., Gao, C., Wang, N., Li, X., Wang, X., Zhu, X., 2015. Comprehensive calibration and validation site for information remote sensing. *ISPRS Archives*, 40(7), 1233.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. *CVPR Workshops*, 136–144.

Marquina, A., Osher, S. J., 2008. Image super-resolution by TV-regularization and Bregman iteration. *Journal of Scientific Computing*, 37(3), 367–382.

Märtens, M., Izzo, D., Krzic, A., Cox, D., 2019. Super-resolution of PROBA-V images using convolutional neural networks. *Astrodynamics*, 3(4), 387–402.

Merino, M. T., Nunez, J., 2007. Super-resolution of remotely sensed images with variable-pixel linear reconstruction. *IEEE TGRS*, 45(5), 1446–1457.

Molini, A. B., Valsesia, D., Fracastoro, G., Magli, E., 2019. DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images. *IEEE TGRS*, 1-13.

Murthy, K., Shearn, M., Smiley, B. D., Chau, A. H., Levine, J., Robinson, M. D., 2014. Skysat-1: very high-resolution imagery from a small satellite. *Sensors, Systems, and Next-Generation Satellites*, 9241, SPIE, 367 – 378.

Nasrollahi, K., Moeslund, T. B., 2014. Super-resolution: a comprehensive survey. *Machine vision and applications*, 25(6), 1423–1468.

Nguyen, N., Milanfar, P., 2000. A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution). *Circuits, Systems and Signal Processing*, 19(4), 321–338.

Peng, Y., Yang, F., Dai, Q., Xu, W., Vetterli, M., 2012. Super-resolution from unregistered aliased images with unknown scalings and shifts. *ICASSP*, 857–860.

Potts, D., Steidl, G., Tasche, M., 2001. *Fast Fourier Transforms for Nonequispaced Data: A Tutorial*. Birkhäuser Boston, Boston, MA, 247–270.

Rais, M., Morel, J.-M., Thiebaut, C., Delvit, J.-M., Facciolo, G., 2016. Improving wavefront sensing with a Shack–Hartmann device. *Applied Optics*, 55(28), 7836.

Reichenbach, S. E., Park, S. K., Narayanswamy, R., 1991. Characterizing digital image acquisition devices. *Optical Engineering*, 30(2), 170–178.

Roland, J. K. M., 2015. A study of slanted-edge MTF stability and repeatability. *Image Quality and System Performance XII*, 9396, SPIE, 181 – 189.

Shannon, C. E., 1948. A mathematical theory of communication. *Bell system technical journal*, 27(3), 379–423.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *CVPR*, 1874–1883.

Strohmer, T., 1995. On discrete band-limited signal extrapolation. *Contemporary Mathematics*, 190, 323–323.

Strohmer, T., 1997. Computationally attractive reconstruction of bandlimited images from irregular samples. *IEEE Transactions on image processing*, 6(4), 540–548.

Sánchez, J., 2016. The Inverse Compositional Algorithm for Parametric Registration. *IPOL*, 6, 212–232.

Takeda, H., Farsiu, S., Milanfar, P., 2007. Kernel regression for image processing and reconstruction. *IEEE Transactions on image processing*, 16(2), 349–366.

Takeda, H., Milanfar, P., Protter, M., Elad, M., 2009. Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing*, 18(9), 1958–1975.

Tekalp, A. M., Ozkan, M. K., Sezan, M. I., 1992. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. *ICASSP*, 3, 169–172.

Tom, B. C., Katsaggelos, A. K., 1995. Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. *ICIP*, IEEE, 539–542.

Vandewalle, P., 2006. Super-resolution from unregistered aliased images. Technical report, EPFL.

Wang, X., Chan, K. C., Yu, K., Dong, C., Change Loy, C., 2019. Edvr: Video restoration with enhanced deformable convolutional networks. *CVPR Workshops*.

Wronski, B., Garcia-Dorado, I., Ernst, M., Kelly, D., Krainin, M., Liang, C.-K., Levoy, M., Milanfar, P., 2019. Handheld multi-frame super-resolution. *ACM TOG*, 38(4), 1–18.

Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L., 2016. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128, 389–408.