Super-Resolution of Multispectral Multiresolution Images from a Single Sensor

Charis Lanaras[1] José Bioucas-Dias[2] Emmanuel Baltsavias[1] Konrad Schindler[1]
[1]Photogrammetry and Remote Sensing, ETH Zürich
[2]Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa
{lanaras,manos,schindler}@geod.baug.ethz.ch, bioucas@lx.it.pt

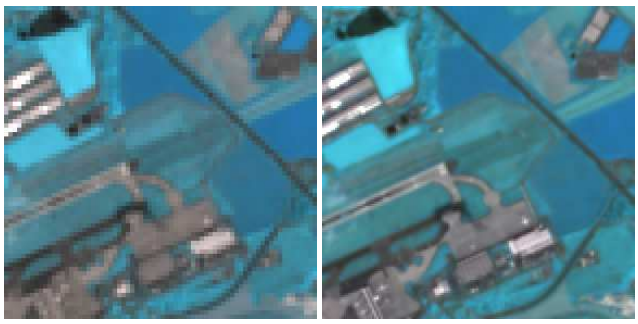
Abstract

Some remote sensing sensors, acquire multispectral images of different spatial resolutions in variable spectral ranges (e.g. Sentinel-2, MODIS). The aim of this research is to infer all the spectral bands, of multiresolution sensors, in the highest available resolution of the sensor. We formulate this problem as a minimisation of a convex objective function with an adaptive (edge-reserving) regulariser. The data-fitting term accounts for individual blur and downsampling per band, while the regulariser "learns" the discontinuities from the higher resolution bands and transfers them to other bands. We also observed that the data can be represented in a lower-dimensional subspace, reducing the dimensionality of the problem and significantly improving its conditioning. In a series of experiments with simulated data, we obtain results that outperform state-of-the-art, while showing competitive qualitative results on real Sentinel-2 data.

1. Introduction

In remote sensing there are a growing number of sensors that acquire multi-spectral images in which the spatial resolution (or Ground Sampling Distance – GSD) varies across different spectral bands. This is mainly the case for satellite sensors and prominent examples are MODIS, ASTER, VIIRS, Worldview-3 and Sentinel-2 (S2). The underlying reason is that design considerations, sensor hardware limitations, and further influences like atmospheric absorption, necessitate the use of different spatial resolutions for various channels, so as to achieve a satifactory signal-to-noise ratio (SNR). It is unlikely that all such resolution differences will go away with hardware improvements, thus it is natural to try and improve the resolution of the coarser bands computationally, by exploiting the structure in the optical signal.

Without loss of generality, we focus here on Sentinel-2, which records 13 bands at 3 different (spatial) resolutions, with good geometric and spectral details. A list of the S2 bands is given in Table 1. The frequent revisit rate, global access and freely available data constitute a great utility for a wide range of applications based on remote sensing. Sentinel-2 data is, among others, used for the monitoring of vegetation, soil, water cover, inland waterways and coastal areas, as well as for the estimation of geophysical variables.



(a) Sentinel-2 input (b) Result of SupReME

Figure 1. Results on real Sentinel-2 images with a false color composite of bands (B5, B6, B7). *Left*: The input 20 m resolution bands. *Right*: The super-resolved image.

The aim of the present work is to increase the spatial resolution of lower-resolution bands, such that all bands have the same, maximal resolution. This is obviously desirable from a product point of view, since then all further down-stream products can be derived at high spatial resolution and can create new opportunities, *e.g.* better observation of clouds/ice/snow and better estimation of climate variables. The environment and climate applications can especially benefit from this approach, since they require many bands in the infrared spectrum.

Interestingly, this problem can be seen as an extension of pan-sharpening, with the main differences that (i) there can be more than one channel at the highest resolution, and (ii) the high-resolution bands need not spectrally overlap the lower-resolution ones. These two conditions of classical pan-sharpening, however, are not met by any of the mentioned instruments. Some recent work has explored an

Table 1. The 13 Sentinel-2 bands.

Band	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Center wavelength [nm]	443	490	560	665	705	740	783	842	865	945	1380	1610	2190
Spatial Resolution [m]	60	10	10	10	20	20	20	10	20	60	60	20	20

intermediate solution, where a single high-resolution band (a "virtual panchromatic image") is synthesised and used as input to a standard pan-sharpening approach. We take one step further and disconnect from the classic pan-sharpening paradigm, integrating channels of different spectral sensitivity and spatial resolution into a compact imaging model.

Contribution. We propose a high-quality solution for this "smart upsampling". Our method relies strongly on the observation model of the imaging (blurring and downsampling) process that generates the low-resolution images. Given the nature of blurring and downsamling, the inversion of the model is an ill-posed problem. To overcome this, we exploit the fact that spectral bands are correlated and can thus be represented in a lower-dimensional subspace, where most of the energy is contained. We learn this subspace from the input data, to reduce the number of unknowns and further stabilise the computation. Moreover, we fully exploit the textural information from the high resolution bands to tailor a regulariser, by encoding the discontinuities of the data and propagating the spatial information to the lower resolution bands. The main underlying assumption is that the discontinuities are likely to be located in the same location across all bands. Given the quadratic data term and the adaptive quadratic regulariser we formulate a convex problem to invert the observation model, and propose an efficient numerical solver. The proposed method performs, in a single step, a super-resolution for all lower resolution bands to the maximum resolution. We term our method *SupReME* (SUPer-REsolution for multispectral Multiresolution Estimation).

2. Related Work

The problem of enhancing the spatial resolution based on spectral bands from the same sensor has been approached for (at least) MODIS, ASTER, and more recently for Sentinel-2 and VIIRS. Tonooka [16] performs super-resolution for the thermal infrared and shortwave infrared bands of ASTER, using the visible and near infrared bands, by means of spectral similarity. Similarly, Aiazzi *et al*. [2] super-resolve the ASTER TIR channels by injecting spatial detail from the VNIR channels using a Generalized Laplacian Pyramid method. Also to super-resolve the ASTER TIR channels, Fasbender *et al*. [7] propose a general Bayesian data fusion approach. Thermal sharpening of VIIRS data was investigated by Picaro *et al*. [10], by considering various "thermal sharpening" methods.

A few studies also increase the resolution of the 500 m bands in MODIS. Sirguey *et al*. [15] use wavelet-based multiresolution analysis, based on the ARSIS concept [11], in an "injection-type" method, which was applied for sub-pixel monitoring of seasonal snow cover [14]. Trishchenko *et al*. [17] proposed a non-linear regression and normalization to preserve radiometric consistency of the super-resolved channels. This method was further applied for producing clear-sky, cloud and cloud shadow masks at 250 m resolution [9]. In technically similar work, Wang *et al*. [20] propose a pan-sharpening method termed ATPRK (area-to-point regression kriging) that also starts from regression modeling, then residual upsampling in order to comply with the spectral properties of the lower-resolution bands.

The same method (ATPRK) has been applied to sharpen the 20 m bands of S2 [21] with the help of the 10 m bands. Since ATPRK only accepts one high resolution image, it is proposed to use either the spectrally nearest 10 m band, or an average of all 10 m bands. Du *et al*. [6] compare four different pan-sharpening methods to sharpen the SWIR band (B11) of S2. Their application is to compute the Modified Normalized Difference Water Index (MNDWI) at 10 m resolution, for monitoring of water bodies. Moreover, Vaiopoulos and Karantzalos [18] compared 21 different fusion algorithms to sharpen the VNIR and SWIR bands of Sentinel-2. Surprisingly, they report that standard bicubic interpolation outperforms sophisticated pan-sharpening methods in terms of the QNR metric (*Quality with No Reference*, [3]). They believe that this is, at least in part, due to a bias in the QNR metric. Most recently, Brodu [4] introduced a so-called "geometry of scene elements" to unmix (super-resolve) the low-resolution bands of S2, while preserving their overall reflectance. This is the only work we are aware of that also enhances the low resolution bands (60 m).

We take a somewhat different approach: We set up a joint model of the physical imaging process in all the different channels, and invert that model with variational calculus.

3. Problem Formulation

Without loss of generality, we present our model for the case of S2. Our inputs have $L_1 = 4$ high resolution bands with GSD 10 m, $L_2 = 6$ medium resolution bands (20 m), and $L_6 = 3$ low resolution bands (60 m), in total $L = L_1 + L_2 + L_6$ spectral bands. The output shall have

10 m resolution for all bands, thus the upsampling factors are $r_1 = 1$, $r_2 = 2$, $r_6 = 6$, and a fixed image area contains $n = n_1 = n/r_1^2$ high-resolution pixels, $n2 = n/r_2^2$ medium-resolution pixels, or $n_6 = n/r_6^2$ low-resolution pixels. To derive our method it is convenient to vectorise the images: the pixel intensities of each individual band are collected into a vector $\mathbf{y}_i$, and the bands are then concatenated (in arbitrary, fixed order) into $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L) \in \mathbb{R}^{L_1 n_1 + L_2 n_2 + L_6 n_6}$. Similarly, the unknown output image is $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L) \in \mathbb{R}^{Ln}$, except that now all bands have the same, maximal, resolution, $\mathbf{x}_i \in \mathbb{R}^n$. Accordingly, the output bands can also be reformatted into a matrix when necessary, $\mathbf{X} = [\mathbf{x}_1^\top; \mathbf{x}_2^\top; \ldots; \mathbf{x}_L^\top] \in \mathbb{R}^{L \times n}$, and we have $\mathbf{x} = \text{vec}(\mathbf{X}^\top)$.

The input and output images are, up to noise, related through the following observation model:

$$\mathbf{y} = \mathbf{MBx} \,, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{(L_1 n_1 + L_2 n_2 + L_6 n_6) \times Ln}$ and $\mathbf{B} \in \mathbb{R}^{Ln \times Ln}$ are two block diagonal matrices. Every sub-block, in both of them, acts on one spectral band. For $\mathbf{M}$, the blocks represent the sampling of $\mathbf{x}$ to obtain $\mathbf{y}$, *i.e.*, blocks are identity matrices for high-resolution channels, and downsampling functions (masks) for other channels, *i.e.*, a subset of rows of the identity matrix. The blur matrix $\mathbf{B}$ is a block-circulant-circulant-block (BCCB) matrix, where each block represents a 2D cyclic convolution, associated with the point spread function (PSF) of the corresponding band at the resolution of $\mathbf{x}$ (highest spatial resolution). The blur can be different in every spectral band[1], and is assumed to be spatially invariant. The model in eq. (1) in its raw form is obviously ill-posed, since there are fewer observations than unknowns.

3.1. Subspace representation

It has often been observed that multi-spectral image data are correlated and can be projected into a subspace of lower dimensionality without losing information. For the 13 bands of S2, we find that $> 99\%$ of the signal energy is retained in the $p = 6$ largest components of the correlation-based eigen-decomposition. The dimensionality reduction significantly reduces the number of unknowns, and proves to be a (implicitly or explicitly used) key ingredient for the envisaged super-resolution. Formally, the columns of $\mathbf{X}$ (*i.e.*, spectral vectors) live in a subspace spanned by the columns of $\mathbf{U} \in \mathbb{R}^{L \times p}$ and thus we may write:

$$\mathbf{X} = \mathbf{UZ} \,, \tag{2}$$

where $\mathbf{Z} \in \mathbb{R}^{p \times n}$ are the representation coefficients with respect to $\mathbf{U}$. We assume that $\mathbf{U}$ is semi-unitary. Vectorising the matrices yields

$$\begin{aligned} \mathbf{x} = \text{vec}(\mathbf{X}^\top) &= \text{vec}(\mathbf{I}\mathbf{Z}^\top\mathbf{U}^\top) \\ &= (\mathbf{U} \otimes \mathbf{I})\text{vec}(\mathbf{Z}^\top) = (\mathbf{U} \otimes \mathbf{I})\mathbf{z} \,, \end{aligned} \tag{3}$$

where $\mathbf{I}$ is an identity matrix with suitable dimensions.

With the dimensionality reduction, the observation model of eq. (1) becomes

$$\mathbf{y} = \mathbf{MB}(\mathbf{U} \otimes \mathbf{I})\mathbf{z} \,. \tag{4}$$

Note that, algebraically, the reduced problem is no longer ill-posed, as long as $pn < L_1 n_1 + L_2 n_2 + L_6 n_6$. However, it is still very ill-conditioned. A direct solution is not practical, as it is extremely sensitive to the presence of noise (even with low magnitude).

Estimation of the subspace. To compute the subspace $\mathbf{U}$, we need to have access to $\mathbf{X}$, respectively to the span($\mathbf{X}$). However, we only have a version of $\mathbf{X}$ that is blurred and downsampled, namely $\mathbf{y}$. Our objective is to estimate the span($\mathbf{X}$), using a blurred version of $\mathbf{X}$, $\mathbf{XQ}$, where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is a blurring matrix representing a 2D cyclic convolution. The main motivation behind this is that the blurring operator $\mathbf{Q}$ does not affect the span($\mathbf{X}$), *i.e.* span($\mathbf{X}$) = span($\mathbf{XQ}$), given that it is a linear combination of the columns of $\mathbf{X}$ and that rank($\mathbf{Q}$) $\geq p$. To do so (i) we upsample all the bands of $\mathbf{y}$ to the same high resolution, using bicubic interpolation and (ii) we blur each band, such that the blur of all the bands is more or less the same, *i.e.* equivalent to the strongest blur. This image serves as a the best approximation of $\mathbf{XQ}$. Next, we perform singular value decomposition analysis on the blurred data. We retain the first p left singular vectors (with decreasing order of singular values) as the columns of $\mathbf{U}$. We assume in this way that the columns of $\mathbf{U}$ span the same subspace as the columns of $\mathbf{X}$.

4. Proposed Solution

To invert our image model, we solve the optimisation

$$\min_{\mathbf{z}} \; \|\mathbf{MB}(\mathbf{U} \otimes \mathbf{I})\mathbf{z} - \mathbf{y}\|^2 + \lambda \phi_{\mathbf{w},\mathbf{q}}(\mathbf{D}_h\mathbf{z}, \mathbf{D}_v\mathbf{z}) \,, \tag{5}$$

where $\phi_{\mathbf{w},\mathbf{q}}$ is a regularisation term, based on weights $\mathbf{w}, \mathbf{q}$ (see Sec. 4.1) and λ is the regularisation strength. $\mathbf{D}_h, \mathbf{D}_v \in \mathbb{R}^{Ln \times Ln}$ are two block-diagonal linear operators (each with identical blocks) that approximate horizontal and vertical derivatives of the images in $\mathbf{z}$. For simplicity, we treat these matrices with periodic boundary conditions as cyclic convolutions. For the regulariser, we choose the quadratic form

$$\phi_{\mathbf{w},\mathbf{q}}(\mathbf{z}) = \sum_{i=1}^{p} \sum_{j=1}^{n} \{q_i w_j (\mathbf{H}_h \mathbf{z}_i)_j^2 + q_i w_j (\mathbf{H}_v \mathbf{z}_i)_j^2\} \,, \tag{6}$$

[1]This is why we use the vector notation $\mathbf{x}$ and $\mathbf{y}$. Matrix format is impractical if one allows for band-specific blur.

where the index i runs over the subspace dimensions and j over all pixels for basis vector i. $\mathbf{H}_h, \mathbf{H}_v \in \mathbb{R}^{n\times n}$ are the individual blocks of the finite difference operators (derivatives) $\mathbf{D}_h$ and $\mathbf{D}_v$, respectively, such that $\mathbf{D}_h = \mathbf{I}\otimes\mathbf{H}_h$ and $\mathbf{D}_v = \mathbf{I}\otimes\mathbf{H}_v$.

To solve the minimisation (eq. 5), we use C-SALSA [1], an instance of ADMM, as follows:

$$\begin{aligned} \min_{\mathbf{z},\mathbf{v_1},\mathbf{v_2},\mathbf{v_3}} \quad & \|\mathbf{MBv_1}-\mathbf{y}\|^2 + \lambda\phi_{\mathbf{w},\mathbf{q}}(\mathbf{v_2},\mathbf{v_3}) \qquad (7)\\ \text{subject to} \quad & \mathbf{v_1} = (\mathbf{U}\otimes\mathbf{I})\mathbf{z} \\ & \mathbf{v_2} = \mathbf{D}_h\mathbf{z} \\ & \mathbf{v_3} = \mathbf{D}_v\mathbf{z}\,. \end{aligned}$$

This splitting makes it possible to solve each individual problem in a much easier way. The Augmented Lagrangian of the above problem is:

$$\begin{aligned} &\mathcal{L}(\mathbf{z},\mathbf{v_1},\mathbf{v_2},\mathbf{v_3},\mathbf{d_1},\mathbf{d_2},\mathbf{d_3}) = \|\mathbf{MBv_1}-\mathbf{y}\|^2 \\ &\quad + \frac{\mu}{2}\|(\mathbf{U}\otimes\mathbf{I})\mathbf{z}-\mathbf{v_1}-\mathbf{d_1}\|^2 + \lambda\phi_{\mathbf{w},\mathbf{q}}(\mathbf{v_2},\mathbf{v_3}) \\ &+ \frac{\mu}{2}\|\mathbf{D}_h\mathbf{z}-\mathbf{v_2}-\mathbf{d_2}\|^2 + \frac{\mu}{2}\|\mathbf{D}_v\mathbf{z}-\mathbf{v_3}-\mathbf{d_3}\|^2\,, \quad (8) \end{aligned}$$

where $\mathbf{d_1}, \mathbf{d_2}$ and $\mathbf{d_3}$ are the scaled Lagrange multipliers and μ is a positive weight. The solution of eq. (8) with respect to $\mathbf{z}$ is

$$\begin{aligned} &\mathbf{z} = (\mathbf{I} + \mathbf{D}_h\mathbf{D}_h^\top + \mathbf{D}_v\mathbf{D}_v^\top)^{-1} \\ &\{(\mathbf{U}\otimes\mathbf{I})(\mathbf{v_1}+\mathbf{d_1}) + \mathbf{D}_h^\top(\mathbf{v_2}+\mathbf{d_2}) + \mathbf{D}_v^\top(\mathbf{v_3}+\mathbf{d_3})\}\,, \end{aligned} \quad (9)$$

given that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$. The matrix to be inverted is block diagonal, from the definition of $\mathbf{D}_h$ and $\mathbf{D}_v$. We can therefore solve separately for each subspace dimension of $\mathbf{z}$. Even so, the numerical inversion of a block element would be impossible on a normal computer in terms of memory and operation count. The trick is to exploit the structure of $\mathbf{D}_h$ and $\mathbf{D}_v$. They are both BCCB matrices and thus, one can compute the solution in the frequency domain [13]. In the following we minimise the Lagrangian (eq. 8) with respect to $\mathbf{v_1}$:

$$\begin{aligned} \mathbf{v_1} = &(\mathbf{B}^\top\mathbf{M}^\top\mathbf{MB} + \mu\mathbf{I})^{-1} \\ &(\mathbf{B}^\top\mathbf{M}^\top\mathbf{y} + \mu(\mathbf{U}\otimes\mathbf{I})\mathbf{z}) - \mathbf{d_1})\,. \end{aligned} \quad (10)$$

Recently, it has been shown that systems of equations involving this type of matrices can be efficiently solved in the frequency domain [22]. The inversion is done independently for each diagonal block, *i.e.*, each dimension of the subspace. The unknown $\mathbf{v_2}$ and $\mathbf{v_3}$ in eq. (8) is obtained from

$$\begin{aligned} &\mathbf{v_2},\mathbf{v_3} \in \arg\min_{\mathbf{v_2},\mathbf{v_3}} \lambda\phi_{\mathbf{w},\mathbf{q}}(\mathbf{v_2},\mathbf{v_3}) \\ &+ \frac{\mu}{2}\|\mathbf{D}_h\mathbf{z}-\mathbf{v_2}-\mathbf{d_2}\|^2 + \frac{\mu}{2}\|\mathbf{D}_v\mathbf{z}-\mathbf{v_3}-\mathbf{d_3}\|^2\,. \end{aligned} \quad (11)$$

Algorithm 1 SupReME. Solver for the optimisation of (5).

Require: data:$\mathbf{y}$, sensor blurs:$\mathbf{B}$, regularisation parameter λ, weigths $\mathbf{q}$
Estimate the subspace $\mathbf{U}$ and weights $\mathbf{w}$ from eq. (14,15)
$k \leftarrow 0$
Initialise $\mathbf{v_1}^{(0)}, \mathbf{v_2}^{(0)}, \mathbf{v_3}^{(0)}, \mathbf{d_1}^{(0)}, \mathbf{d_2}^{(0)}$ and $\mathbf{d_3}^{(0)}$
while *not converged* **do**
  $k \leftarrow k+1$
  Estimate $\mathbf{z}^{(k)}$ with eq. (9)
  Estimate $\mathbf{v_1}^{(k)}, \mathbf{v_2}^{(k)}, \mathbf{v_3}^{(k)}$ with eq. (10,12)
  Estimate $\mathbf{d_1}^{(k)}, \mathbf{d_2}^{(k)}, \mathbf{d_3}^{(k)}$ with eq. (13)
end while
return $\mathbf{x} = (\mathbf{U}\otimes\mathbf{I})\mathbf{z}$

The solution is uncoupled with respect to $\mathbf{v_2}$ and $\mathbf{v_3}$ and can be computed for each element as

$$\mathbf{v_2}_{ij} = \frac{\mu(\mathbf{H}_h\mathbf{z}_i - \mathbf{d_2}_i)_j}{\mu + 2\lambda w_i q_j}, \quad \mathbf{v_3}_{ij} = \frac{\mu(\mathbf{H}_v\mathbf{z}_i - \mathbf{d_3}_i)_j}{\mu + 2\lambda w_i q_j}\,, \quad (12)$$

where i represents the ith image in $\mathbf{z}$ and j each pixel.

Finally, we update the Lagrange multipliers as:

$$\begin{aligned} \mathbf{d_1}^{(k)} &= \mathbf{d_1}^{(k-1)} - ((\mathbf{U}\otimes\mathbf{I})\mathbf{z} - \mathbf{v_1}) \\ \mathbf{d_2}^{(k)} &= \mathbf{d_2}^{(k-1)} - (\mathbf{D}_h\mathbf{z} - \mathbf{v_2}) \\ \mathbf{d_3}^{(k)} &= \mathbf{d_3}^{(k-1)} - (\mathbf{D}_v\mathbf{z} - \mathbf{v_3})\,. \end{aligned} \quad (13)$$

The complete optimization scheme is summarized in Algorithm 1. Initial values for the parameters $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{d_1}$, $\mathbf{d_2}$ and $\mathbf{d_3}$ can be chosen arbitrarily, since the problem is convex. The convergence to the global minimum is guaranteed for any $\mu > 0$, see [1].

4.1. Adapting the spatial regularisation

For the quadratic smoothing used as spatial regulariser, we can introduce weights $\mathbf{w}$ for each pixel, so as to reduce smoothing across discontinuities. We take

$$\mathbf{w} = \exp(-\frac{\mathbf{g}^2_{max}}{2\sigma_s}) \quad (14)$$

and define $\mathbf{g}_{max} = \max(\mathbf{g}(\mathbf{x}_1), \ldots, \mathbf{g}(\mathbf{x}_{L_1}))$, of the high resolution bands, where the $\max(\cdot)$ operator is applied element-wise (per pixel), $\mathbf{g}(\cdot)$ is the Prewitt image gradient magnitude. The weights (eq. 14) are in the range $[0\ldots1]$. We found that very strong edges are downweighted too strongly, and truncate values smaller than 0.5:

$$\mathbf{w} = \max(0.5, \mathbf{w})\,. \quad (15)$$

Empirically, this improves the result.

By their definition, the subspace coefficients have different value ranges. The first few dimensions (basis vectors)

cover most of the energy (information), and thus are – relatively – less affected by noise, whereas in the last dimensions the noise is more dominant. It is therefore necessary to apply progressively stronger regularisation. We simply reweight the regulariser with $\mathbf{q} \in \mathbb{R}^p$, in an heuristic fashion. At this stage the best setting is

$$\mathbf{q} = [1\ 1.5\ 4\ 8\ 15\ 15\ 20]\ . \tag{16}$$

We tried using principal component based schemes, but it was not leading to better results compared to the heuristic approach, that performed well across all tested datasets.

4.2. Implementation details

Our method requires the PSF of each band. The variance of the Gaussian blur (termed sdf) included in $\mathbf{B}$ differs for each band, and is computed from the calibrated MTF (Modulation Transfer Function) supplied by ESA as part of the meta-data [5]:

$$\text{sdf} = r_i \sqrt{\frac{-2\ln(\text{mtf})}{\pi^2}}\ , \tag{17}$$

with $i = 1, 2, 6$ for the different resolutions. The highest resolution bands are assumed to have a point response (sdf $= 0$). Empirically, the estimation is not very sensitive to small variations of the blur. For better numerical conditioning we also normalize the images before the processing, such that their mean squared intensities are 1 (normalisation has been undone for the qualitative results below). As mentioned before, the blur $\mathbf{B}$ and the difference matrices $\mathbf{D}_h$ and $\mathbf{D}_v$ have periodic boundary conditions. For this reason we remove a border of 18 pixels, to suppress artifacts originating from the periodic boundaries. The subspace dimension is set to $p = 7$, as it was able to capture the spectral variability in all our experiments. The value $\mu = 0.2$ is kept constant for the whole processing, and the normalization of the weights $\mathbf{w}$ is set to $\sigma_s = 1$.

As convergence criterion for the optimiser, we iterate until the residuals associated with the variable splitting fall bellow 0.001, or at most 200 iterations. With this criterion we have obtained a solution for every scenario we tested. For an image of dimensions 180×180 pixels the program runs on average less than 10 seconds on a Intel Xeon E5 3.2 GHz CPU, in a MATLAB implementation.

5. Experimental Results

Quality indices. As a primary quality metric for the evaluation we use the signal-to-reconstruction error (SRE), given in dB. It is computed from the reconstructed image $\hat{\mathbf{x}}$ and the ground truth $\mathbf{x}$ as:

$$\text{SRE} = 10\log_{10}\frac{\mu_{\mathbf{x}}^2}{\|\mathbf{x}-\hat{\mathbf{x}}\|^2/n}\ , \tag{18}$$

where $\mu_{\mathbf{x}}$ is the average value of $\mathbf{x}$. As a complementary metric we also report the Spectral Angle Mapper (SAM) [23] for each image, defined as

$$\text{SAM} = \frac{1}{n(L_2+L_6)}\sum \arccos\frac{\mathbf{X}_{:j}^\top \hat{\mathbf{X}}_{:j}}{\|\mathbf{X}_{:j}\|_2\|\hat{\mathbf{X}}_{:j}\|_2}\ , \tag{19}$$

where $\mathbf{X}_{:j}$, $\hat{\mathbf{X}}_{:j}$ are the spectral values of the L_2 and L_6 bands at pixel j, of the ground truth and reconstruction respectively. The SAM is computed only for the super-resolved bands (per pixel) and is averaged over the whole image. It is given in degrees.

The QNR value has also been used to judge super-resolution results [18]. However, as mentioned in Sec. 2 the result can be misleading. The QNR was primarily designed to evaluate pan-sharpening results and appears unsuitable to compare conceptually different types of algorithms [19].

Baselines. As baselines to compare against, we use the following 4 methods. As the simplest solution to the problem we use a bicubic upsampling, implemented in MATLAB. The second baseline is MTF-GLP-HPM-PP, which is an injection-based pan-sharpening method, initially presented by Lee and Lee [8]. That method performed best against other pan-sharpening methods in [18] and we follow the same experimental procedure as described there. The third baseline, termed SuperRes, is the method of [4]. The last baseline is ARTPK [21], which reported the best results for our task that we are aware of. From the two variants described in that publication ("selected" and "synthesized"), we compare against "synthesized", which is the average of all high resolution bands and gave the best numerical results. For all baselines, code has been made available, and we use those original implementations.

5.1. Simulated data

For the simulated data, we use the Open Science Dataset of APEX [12], available online[2]. APEX is a hyperspectral sensor that covers the wavelengths from 0.4 to 2.5 μm and can be used to simulate S2 images, given the nominal spectral response of S2 (Table 1). We took care to model the spectral downgrading as well as possible, by using a few more bands, if the inputs were noisy. From the APEX scenes, we created two sets of inputs with different resolutions, and consequently different levels of object detail. The first set has maximum resolution ≈2m, which is the nominal resolution of the APEX image. The second one has maximum resolution ≈10m, for a strict simulation of S2. The corresponding medium and low resolution images are 2 and 6 times degraded respectively, by pixel aggregation with the known PSF of S2. Band B10 is discarded from all

[2] http://www.apex-esa.org/content/free-data-cubes

Table 2. Numerical results on simulated data, SRE in dB and SAM (last column) in degrees. The input APEX data are in their original resolution (2 m). Note that, the resolution mentioned does not correspond with the actual resolution of the image, but gives an overview of the resolution of the S2 bands. Best results in bold font.

	B1	B5	B6	B7	B8a	B9	B11	B12	
Method	60m	20m	20m	20m	20m	60m	20m	20m	SAM
Bicubic upsampling	8.39	14.97	15.88	15.86	15.93	12.66	19.55	17.12	7.80
MTF-GLP-HPM-PP	–	19.38	21.66	21.48	22.42	–	–	–	–
SuperRes	12.13	19.58	20.40	20.48	14.00	**21.99**	20.78	18.59	7.84
ARTPK	–	**25.29**	**29.80**	30.35	31.08	–	20.93	20.45	–
SupReME	**22.01**	25.16	27.87	**31.29**	**32.12**	19.78	**21.42**	**21.16**	**4.63**

Table 3. Numerical results on simulated data, SRE in dB and SAM (last column) in degrees. The input APEX data has been downsampled to match the true GSD of S2 bands. Best results in bold font.

	B1	B5	B6	B7	B8a	B9	B11	B12	
Method	60m	20m	20m	20m	20m	60m	20m	20m	SAM
Bicubic upsampling	8.35	16.46	17.59	17.51	17.54	13.20	16.70	13.47	5.77
MTF-GLP-HPM-PP	–	2.53	23.37	18.71	23.92	–	–	–	–
SuperRes	11.78	21.22	22.37	22.37	17.54	22.23	17.99	14.59	4.55
ARTPK	–	25.47	**30.84**	31.34	31.84	–	**21.09**	**17.14**	–
SupReME	**20.64**	**26.36**	29.85	**34.26**	**34.23**	**23.69**	19.95	16.41	**2.71**

further processing. That band is primarily there to detect Cirrus clouds, but it does not show any ground structures, and is always very noisy (both in real S2 data and when created from APEX). The dimensions of the image used are in both cases 180×180. Gaussian noise of SNR $= 40$dB has been added to all simulated images. In the current version of SupReME we use $\lambda = 0.005$ for APEX.

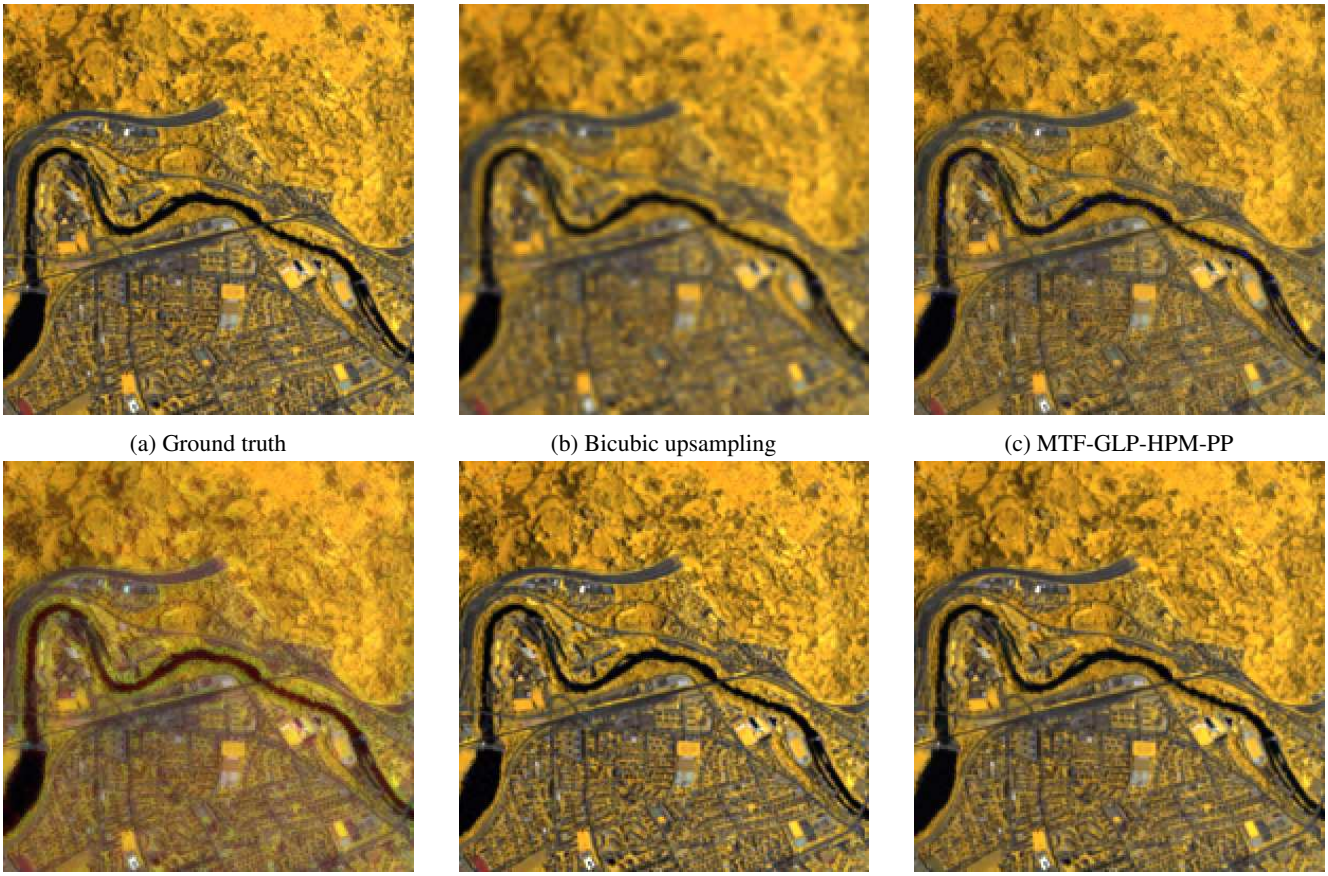
Tabulated numerical results for 2 m and 10 m super-resolution are presented in Tables 2 and 3, respectively. The best results are marked in bold font. ARTPK and SupReME perform best, with SupReME dominating in bands B1, B7, and B8a and ARTPK performing well in B6. Most challenging of all are bands B11 and B12, which lie in the SWIR range, far away from all high-resolution bands. B1 and B9 also have slightly weaker results, simply because the original (input) bands are of low resolution. To be fair, SAM is reported in Tables 2 and 3 for methods that reconstruct all channels. Here, SupReME always delivers the best result. That means that the errors are distributed more evenly across the 8 bands compared to other methods. This is especially useful for applications where spectral properties of materials (their proportions) are important. To compare to ARTPK in terms of SAM, we only evaluate the spectral angle on the 6 bands with 20 m resolution. In both cases SupReME is considerably better and achieves a smaller spectral angle: in the full resolution APEX (2 m), 3.30 *vs* 4.43 and in the reduced resolution APEX (10 m), 2.31 *vs* 2.98.

Figure 2 visually compares the results of all baselines to the 10 m ground truth, using a false color composite of bands (B8a, B6, B5). As expected, Bicubic upsampling and SuperRes lack high-frequency detail and appear blurry. MTF-GLP-HPM-PP is better, but exhibits a number of artifacts. ARTPK and SupReME are clearly best and are visually almost indistinguishable from the ground truth. Figure 3 compares the two top-performing methods ARTPK and SupReME to the 2 m ground truth on a scene detail, using a composite of bands (B12, B11, B7). SupReME in this case gives the best result, whereas ARTPK introduces artifacts at spectral discontinuities, *c.f.* the stadium tartan or the grass area near the top left corner. Notice that both reconstruction results are significantly sharper than the ground truth, likely due to the imaging system of APEX. Its VNIR and SWIR sensors have rather different PSF. In contrast, super-resolution transfers high-resolution detail from spectral bands in different wavelengths, which are sharp. In that sense, computational super-resolution can also be seen as a way to remedy wavelength-dependent blur in imaging systems that record all pixels at the same nominal resolution, by exploiting correlations across spectral bands.

5.2. Real Sentinel-2 data

Complementary to simulated data, we use real Sentinel-2a data, acquired on 29th of April 2016 over the Czech Republic in the vicinity of Prague. In this work we use a cut-out of 180×180 pixels that contains a variety of scene elements, close to the Prague airport, see Figure 4. Since we have no access to the true high-resolution information, we cannot perform any quantitative evaluation.

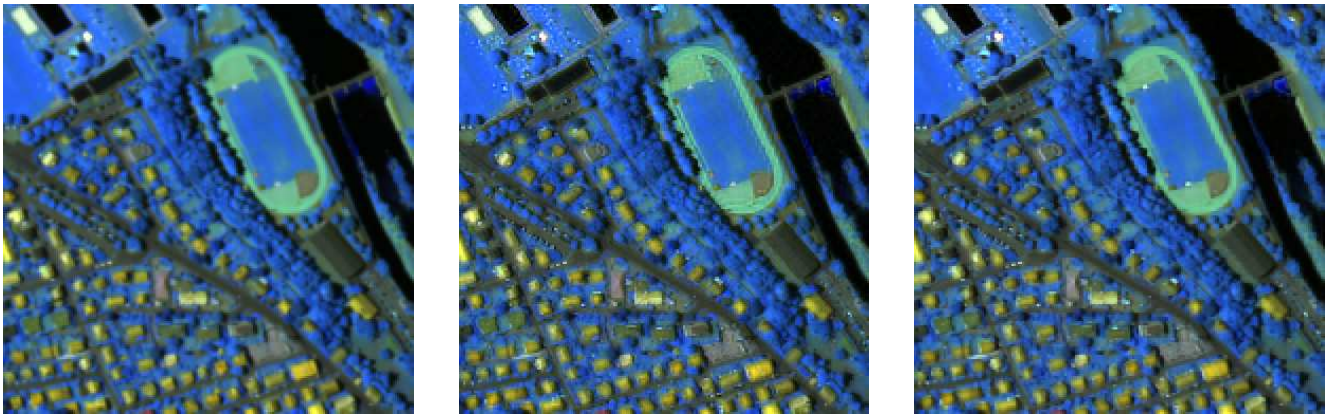
For this data we use a stronger regularisation $\lambda = 0.01$.



(a) Ground truth (b) Bicubic upsampling (c) MTF-GLP-HPM-PP

(d) SuperRes (e) ARTPK (f) SupReME

Figure 2. Comparison of all 5 methods to the ground truth (APEX with 10 m resolution). The false color composite is created with bands (B8a, B6, B5).



(a) Ground truth (b) ARTPK (c) SupReME

Figure 3. Comparison of ARTPK and SupReME to the ground truth (APEX with 2 m resolution). The false color composite is created with bands (B12, B11, B7).

(a) Natural color image of the scene (b) Input low resolution image (c) Super-resolved image with SupReME

Figure 4. Qualitative results on real S2 images. *Left*: The scene in true colors (10 m GSD). *Centre and Right*: A color composite of bands (B12, B9, B11) before and after the super-resolution.

The visual results in Figure 1 and 4 show two different false color composites, with bands (B5, B6, B7) and (B12, B9, B11). In both cases the result of SupReME is compared to the input lower resolution images, resized with nearest neighbor to be able to overlay them with equal pixel size. One clearly sees that, also for real S2 data, the result is much sharper than the input, and does not seem to suffer from any visible spectral (color) distortion.

6. Discussion

So far, work on S2 super-resolution has largely disregarded the potential of the 60 m channels. Even though these bands are primarily intended for atmospheric corrections, we find that they can be upsampled to 10 m, with an accuracy similar to that of the 20 m bands. Although upsampling 1 pixel to 36 may at first sight seem unrealistic, the results indicate that the spectral correlations are strong enough to do it, and these bands could be used to derive high-resolution information, rather than only for calibrating the other high-resolution bands.

We have observed that our approach is not independent of the texture scale (respectively, object size) in the input image. Thus, its performance might change across different sensor resolutions, or across scenes with very different object scales, even if recorded with the same sensor. In future work it should therefore be investigated how well our method generalises to much lower image resolutions (*e.g*. MODIS). Such images have much smaller homogeneous areas and correspondingly more high-frequency texture. We note that, also for such low-resolution, global imagery, super-resolution may have interesting uses. *E.g*., enhancing the even lower resolution of some spectral bands to 250 m could make quite a difference for coastlines and coastal mapping.

In our current setup, we have not attempted to automatically set the parameters p, λ and $\mathbf{q}$. Fixed, empirically chosen values gave good results, but for larger areas of interest it may become necessary to chose them automatically, or even to adaptively change the subspace to best represent the local spectral distribution.

7. Conclusions

We have presented a novel method to deal with the different resolutions in contemporary multi-spectral satellite sensors. Our method integrates information from all channels, independent of their input resolution, into one joint, convex optimization. The model is computationally efficient, and super-resolves all lower-resolution channels to the highest available resolution, by transferring high-frequency information from multiple, non-overlapping high-resolution bands to the entire data cube. The method has been tested on both simulated data and real Sentinel-2 imagery, and achieves state-of-the-art results.

Future extensions shall include locally adaptive subspace projection to deal with strongly varying, locally heterogeneous scene content. Eventually, we aim to super-resolve entire S2 tiles (100×100 km^2). Moreover, as longer time series become available, we would like to investigate high-resolution change detection. It is an interesting open question how to best integrate time series analysis with super-resolution.

Acknowledgements

The authors acknowledge financial support from the Swiss National Science Foundation (SNSF), project no. 200021_162998 and from Portuguese Fundação para a Ciência e Tecnologia (FCT) under the grants UID/EEA/5008/2013 and ERANETMED/0001/2014.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010.

[2] B. Aiazzi, L. Alparone, S. Baronti, L. Santurri, and M. Selva. Spatial resolution enhancement of ASTER thermal bands. In *Proc. SPIE*, volume 5982, pages 59821G–59821G–10, 2005.

[3] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetric Engineering & Remote Sensing*, 74(2):193–200, 2008.

[4] N. Brodu. Super-resolving multiresolution images with band-independant geometry of multispectral pixels. *arXiv preprint arXiv:1609.07986*, 2016.

[5] S. Clerc and MPC Team. S2 MPC - Data Quality Report. [Online: `http://earth.esa.int/documents/247904/685211/Sentinel-2-Data-Quality-Report`, accessed 06 Feb. 2017], 2017.

[6] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li. Water bodies mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 8(4):354, 2016.

[7] D. Fasbender, D. Tuia, P. Bogaert, and M. Kanevski. Support-based implementation of Bayesian data fusion for spatial enhancement: Applications to ASTER thermal images. *IEEE Geoscience and Remote Sensing Letters*, 5(4):598–602, 2008.

[8] J. Lee and C. Lee. Fast and efficient panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):155–163, 2010.

[9] Y. Luo, A. P. Trishchenko, and K. V. Khlopenkov. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment*, 112(12):4167 – 4185, 2008.

[10] G. Picaro, P. Addesso, R. Restaino, G. Vivone, D. Picone, and M. Dalla Mura. Thermal sharpening of VIIRS data. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 7260–7263, 2016.

[11] T. Ranchin and L. Wald. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing*, 66(1):49–61, 2000.

[12] M. E. Schaepman, M. Jehle, A. Hueni, P. D'Odorico, A. Damm, J. Weyermann, et al. Advanced radiometry measurements and earth science applications with the airborne prism experiment (APEX). *Remote Sensing of Environment*, 158:207–219, 2015.

[13] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.

[14] P. Sirguey, R. Mathieu, and Y. Arnaud. Subpixel monitoring of the seasonal snow cover with MODIS at 250 m spatial resolution in the southern Alps of New Zealand: Methodology and accuracy assessment. *Remote Sensing of Environment*, 113(1):160–181, 2009.

[15] P. Sirguey, R. Mathieu, Y. Arnaud, M. M. Khan, and J. Chanussot. Improving MODIS spatial resolution for snow mapping using wavelet fusion and ARSIS concept. *IEEE Geoscience and Remote Sensing Letters*, 5(1):78–82, 2008.

[16] H. Tonooka. Resolution enhancement of ASTER shortwave and thermal infrared bands based on spectral similarity. In *Proc. SPIE*, volume 5657, pages 9–19, 2005.

[17] A. P. Trishchenko, Y. Luo, and K. V. Khlopenkov. A method for downscaling MODIS land channels to 250-m spatial resolution using adaptive regression and normalization. In *Proc. SPIE*, volume 6366, pages 636607–636607–8, 2006.

[18] A. Vaiopoulos and K. Karantzalos. Pansharpening on the narrow VNIR and SWIR spectral bands of Sentinel-2. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B7:723–730, 2016.

[19] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2015.

[20] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao. Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, 166:191 – 204, 2015.

[21] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson. Fusion of Sentinel-2 images. *Remote Sensing of Environment*, 187:241–252, 2016.

[22] Q. Wei, N. Dobigeon, J.-Y. Tourneret, J. Bioucas-Dias, and S. Godsill. R-fuse: Robust fast fusion of multiband images based on solving a Sylvester equation. *IEEE Signal Processing Letters*, 23(11):1632–1636, 2016.

[23] R. H. Yuhas, A. F. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, volume 1, pages 147–149, 1992.