REVIEW

Single-Frame Super-Resolution in Remote Sensing: a Practical Overview

Ruben Fernandez-Beltran[a], Pedro Latorre-Carmona[a*], Filiberto Pla[a]

[a] *Institute of New Imaging Technologies, Universitat Jaume I, SPAIN*

(*November 2016*)

Image acquisition technology is improving very fast from a performance point of view. However, there are physical restrictions that can only be solved using software processing strategies. This is particularly true in the case of super resolution methodologies. Super-resolution techniques have found a fertile application field in airborne and space optical acquisition platforms. Single-frame super resolution methods may be advantageous for some remote sensing platforms and acquisition time conditions. The contributions of this paper are basically two: (1) to present an overview of single-frame super resolution methods, making a comparative analysis of their performance in different and challenging remote sensing scenarios, and (2) to propose a new single-frame super-resolution taxonomy, and a common validation strategy. Finally, we should emphasize that, on the one hand, this is the first time, to the best of our knowledge, that such a review and analysis of single super resolution methods is made in the framework of remote sensing, and, on the other hand, that the new single-frame super-resolution taxonomy is aimed at shedding some light when classifying some types of single-frame super-resolution methods.

Keywords: Super-Resolution; Remote Sensing; Image Reconstruction; Image Learning; Image Quality Assessment

1. Introduction

From early years, improving image resolution has been one of the important concerns of the computer vision and remote sensing research communities in order to deal with acquisition technology limitations. Despite the fact that image sensors are constantly upgraded, there are several physical constraints which can be only addressed by signal and image processing algorithms. Decreasing the physical pixel size beyond a specific limit reduces the incoming light and generates more diffraction effects in the sensor what eventually reduces the final image quality. Furthermore, hardware costs are usually important constraints in many applications because acquisition technology cost and complexity depends on sensor resolution. Therefore, algorithmic-based approaches are sometimes more suitable because they allow improving the image resolution beyond the sensor limits whatsoever.

*Corresponding author. Email: latorre@uji.es

1.1.   *Super-Resolution in Remote Sensing*

The general concept of Super-Resolution (SR) refers to those algorithms aimed at increasing the image resolution, that is, increasing the number of pixels but providing fine details in the resulting image as if a sensor with a higher nominal resolution would have been used. The increasing demand of aerial and satellite imagery applications motivates the development of new data products in order to cope with the new challenges that are constantly appearing nowadays (Bioucas-Dias et al. 2013; Li et al. 2015b,a). Specifically, many of the most important commercial satellites are currently working on providing new super-resolved products of data level-2, i.e. images beyond the sensor limit resolution including all the level-0 pre-processing as well as the radiometric and geometric corrections involved in level-1 data processing.

Depending on the number of input images, two kinds of SR schemes can be distinguished: single-frame and multi-frame. In the former type, only a single image of the objective scene is used to obtain the super-resolved output. In the latter type, several images of the same scene acquired at different positions are used to generate the higher resolution result.

Multi-frame super-resolution methods have been widely applied in hyper-spectral remote sensing, without or in combination with the multi-angular acquisition capability some of these sensors have (Bioucas-Dias et al. 2013; Simoes et al. 2015). For years, the data redundancy inherent in multiple images has caught the research community attention to enhance aerial imagery. However, there are many applications where the existence of multiple images of the same scene is not possible or difficult to obtain. There are two main situations in remote sensing where single-frame super resolution methods can be advantageous in relation to the multi-frame SR counterpart:

- In remote sensing airborne missions that use small platforms with low resolution cheap sensors, and where an on-board image processing may also be an attractive possibility. These low resolution sensors usually have a low spatial resolution capability associated to them.
- When super resolution is applied to satellites that have a long revisit time (for a particular application). The revisit time is the time that a satellite takes to make two consecutive observations of the same point on earth. According to the *Committee of Earth Observations Satellites (CEOS)* webpage (*http://database.eohandbook.com/database/missiontable.aspx*), 49 out of the 140 satellites listed in that database has a revisit period of 7 days or less. A revisit time of one week or more may be too long for some biophysical processes in agriculture, for instance, during the peak growth period (Moreno 2004).

For these reasons, several works in the literature (Loncan et al. 2015; Chavez-Roman and V. 2014; Demirel and G. 2011) advocate the use of single-frame super-resolution in remote sensing.

1.2.   *Limitations in the current literature*

We can find several papers in the literature that provide a good overview of the existing SR algorithms, however few ones are specially aimed at remote sensing and most of them only consider a rather limited number of methods. For instance, a good up-to-date overview is presented in (Nasrollahi and B. 2014) where most

of the papers published on the SR topic up to 2012 are reviewed. Even though the high quality of the work, the lack of an experimental part together with the generality of the application domain makes it difficult to draw practical conclusions in the remote sensing field. Other quality works, like (Yang, Ma, and Yang 2014), provide a comprehensive experimental part but the description and categorisation of methods sound unsatisfactory. Besides, experiments are conducted using only standard test images what makes it difficult to extrapolate conclusions to remote sensing. Other papers, like (Suganya, Mohanapriya, and Vanitha 2013; Zhang et al. 2014; Ramji, Punniakodi, and Praveen 2013), are only focused on super-resolving aerial imaging, however they show some shortcomings especially when it comes to single-frame SR. In all these works, the number of tested methods and the image diversity is rather reduced and often constrained just to multi-frame super-resolution.

In a sense, the lack of aerial imagery benchmarking and standardized implementations of SR algorithms make it difficult to draw practical conclusions for remote sensing. Some important questions like, what are the state-of-the-art SR techniques? or, what are the practical differences among methods? remain still unclear in remote sensing. On the one hand, single-frame SR surveys tackle the SR problem considering a general application despite the fact that the SR performance highly depends on each specific field. On the other hand, remote sensing works tend to neglect single-frame SR methodologies even though these techniques are highly useful in remote sensing (Bioucas-Dias et al. 2013; Yang et al. 2015).

The evaluation criteria is another discussion point in SR (Reibman, Bell, and Gray 2006). There is not an universal criteria to assess the quality of the resulting images and this makes comparison among algorithms difficult. Some works use an experimental validation protocol based on ground-reference images and others include non-reference metrics. This variation of protocols and metrics makes the cross-comparison among SR methods cumbersome, especially in the remote sensing field where both spatial and spectral information are important.

1.3. *Work Objectives and Paper Structure*

Given the above described scenario about, the present work has the following aims: (i) to study the rationale behind many of the state-of-the-art single-frame SR techniques, (ii) to present a clear taxonomy to simplify the categorisation and validation of single-frame SR methods, (iii) to conduct a comprehensive comparison of single-frame SR techniques in remote sensing and (iv) to provide access to a set of algorithms to the research community in order to make benchmarking and evaluation easier in the future. In a sense, this paper pretends to shed light on the single-frame SR research specially applied to remote sensing field.

The rest of the paper is organized as follows. Section 2 presents the background of the work by introducing the concept of SR in remote sensing. Section 3 describes the taxonomy used to classify all the considered methods. Section 4 defines the validation protocol proposed to perform the quality assessment. Section 5 shows and discusses the obtained SR results. Finally, Section 6 draws the main conclusions arisen from the work.

2. Image Super-Resolution to Remote Sensing

2.1. *Imaging Model*

In general, the image acquisition process can be understood as a sequential set of degradations in which the continuous scene signal is converted by the sensor to a discrete output image. In the context of single-frame SR, the imaging model is typically defined as the concatenation of three operators: (i) Blurring ($\mathcal{B}$), which represents the blurring factor introduced by the optical system, i.e. a low pass filter over the continuous scene, (ii) Decimation ($\mathcal{D}$), that sub-samples the blurred continuous space according to the resolution of the sensor, i.e. down-sampling the high resolution scene, and (iii) Noise ($\mathcal{N}$), which introduces a final perturbation to the output image.

Equation (1) shows the imaging model considering the aforementioned operators. The starting point is the High-Resolution (HR) image $\mathbf{I}_{HR}$ corresponding to the continuous scene. Note that the term *continuous scene* conveys the image at the highest possible resolution. Then, $\mathbf{I}_{HR}$ is affected by all the operators in an orderly way to generate the final observed Low-Resolution (LR) image $\mathbf{I}_{LR}$.

$$\mathbf{I}_{LR} = \underbrace{\mathcal{D}}_{decimation} (\underbrace{\mathcal{B}}_{blurring} (\mathbf{I}_{HR})) + \underbrace{\mathcal{N}}_{noise} \tag{1}$$

In the field of remote sensing, the blur operator is specified in the technical description of the instrument as the Point Spread Function (PSF). This function can be seen as a blurring mask which describes how a point of the HR domain is blurred in the LR space during the acquisition process. The theoretical form of the blur, corresponding to a diffraction-limited system, is given by a Bessel function. However, due to some effects like lens aberration or atmospheric turbulence, the blur is often adequately represented by a Gaussian distribution. In other words, it is broadly accepted to approximate the PSF by a Gaussian function whose parameters (μ, σ) depend on the acquisition instrument characteristics. Another important parameter appearing in the sensor technical specifications is the noise level and type. In this case, it is common to consider an additive white Gaussian noise what eventually makes that both PSF and noise are usually approximated by Gaussian functions.

2.2. *Super-Resolution as an inverse problem*

The objective of SR consists in inverting the degradation process generated by the imaging model in order to obtain a super-resolved high resolution output image ($\mathbf{I}_{SR}$) from its corresponding low resolution image ($\mathbf{I}_{LR}$). This inverse problem has an ill-posed nature because different HR images might obtain the same LR image over a specific imaging model.

2.3. *Super-Resolution vs. Interpolation*

It is noteworthy that some authors refer to interpolation algorithms as SR even though there are differences between both concepts. Let us clarify this point in these lines. Interpolation methods project the initial LR image onto an HR *grid* and the missing pixel values are estimated using an interpolation function. Although some

of these functions effectively reduce the aliasing effect caused by the decimation operator of the imaging model, they are not able to recover high-frequency details unlike SR methods.

2.4. *Super-Resolution vs. Pansharpening*

As it has been commented before, SR surveys hardly ever aim at remote sensing imagery but at global applications. However, other remote sensing related topics have been widely addressed in the literature and they can help us to adapt the SR approach to remote sensing. This is the case of Pansharpening techniques. In a nutshell, Pansharpening is a fusion problem in which a high spatial resolution Panchromatic image (PAN) is combined with a lower spatial resolution Multi-Spectral image (MS) in order to obtain a final result which integrates the spatial resolution of PAN and the spectral resolution of MS. Although both Pansharpening and SR have features in common, it is important to highlight their differences. The Pansharpening objective lies on reconstructing just the spectral information of the PAN image because all spatial details are available. Nonetheless, SR is a more general problem in which both spatial and spectral information have to be estimated from the LR image.

2.5. *Super-Resolution of Multi-Spectral images*

The standard SR algorithms are oriented to super-resolving gray scale images, that is, they are focused on enhancing LR images with a single band. In the case of RGB images, the typical super-resolving procedure is based on the transformation into the uniform color space $\mathbf{YC_bC_r}$, where $\mathbf{Y}$ represents the luminance component, $\mathbf{C_b}$ the blue-difference chroma component and $\mathbf{C_r}$ the red-difference chroma component. Figure 1 shows the $\mathbf{YC_bC_r}$ decomposition of an example RGB image. Input multi-spectral images are initially converted to the $\mathbf{YC_bC_r}$ color space. Then, the luminance channel $\mathbf{Y}$ is super-resolved and both $\mathbf{C_b}$ and $\mathbf{C_r}$ bands are interpolated to the target resolution. Finally, the inverse $\mathbf{YC_bC_r}$ transformation is used to generate the final super-resolved output.

Figure 1.: From left to right: RGB image, $\mathbf{Y}$ (luminance), $\mathbf{C_b}$ (blue-difference) and $\mathbf{C_r}$ (red-difference) channels.

Satellites of general purpose like *Geoeye-1* or *IKONOS* usually provide two kinds of images: (i) PAN, a wide spectral bandwidth filtered monochromatic and high resolution image and (ii) MS, a lower resolution multi-spectral image. The final user has to decide whether to super-resolve the PAN image, the MS image or both. In the case of super-resolving the PAN image, SR methods can be directly used because PAN images are made up of a single-band. In the case of MS images, the SR algorithms have to be applied to the luminance component once a false color RGB has been created.

3. A Simplified Taxonomy of Single-Frame Super-Resolution Algorithms

There are different survey papers in the literature (Suganya, Mohanapriya, and Vanitha 2013; Nasrollahi and B. 2014; Yang, Ma, and Yang 2014) that classify single-frame super-resolution methods according to multiple criteria. For instance, the classification criteria may depend on the domain where the SR is performed (e.g. frequency domain or spatial domain), the theoretical tools used to cope with the SR problem (e.g. Gradient Profiles, Neighbourhood Embeddings, Neural Networks, Bayesian Mixture Models, etc.) or even on the knowledge of the process parameters (e.g. blind or non-blind SR). However, many SR methods gather features of several of those groups which eventually makes the definition of a particular algorithm in terms of current taxonomies difficult to be proposed.

In this paper, we propose a relaxed taxonomy based on the functional properties of SR methods inspired by the taxonomy usually applied in Pansharpening (Vivone et al. 2015). Therefore, instead of using a main division of methods based on particular features about how they implement their solutions, we would rather consider the functional definition of the rationale behind the algorithms considered in this research. The following sections describe the three considered SR groups in detail: (i) reconstruction-based methods, (ii) learning based methods and (iii) hybrid methods.

3.1. *Super-Resolution based on Image Reconstruction*

The SR methods based on image REconstruction (RE) pursue to reproduce features appearing in the LR image to a higher resolution level. In other words, they are aimed at obtaining a final super-resolved image with the same perceptual properties of the LR image but in a higher scale. Somehow, RE methods try to replicate the quality of the LR domain after an initial interpolation in order to avoid the blurring and aliasing effects caused by interpolation algorithms.

Figure 2 shows the general scheme related to RE methods. Specifically, this sort of methods are made up of three main steps. In the first one (stage 1), the LR image ($\mathbf{I}_{\mathrm{LR}}$) is upscaled to the target resolution using a regular interpolation algorithm, such us bicubic or lanczos interpolation. The second stage is based on extracting certain physical properties from $\mathbf{I}_{\mathrm{LR}}$ to figure out the quality of the details in the LR domain. Finally, in the third step (stage 3) both the interpolated image $\mathbf{I}_{\mathrm{LRI}}$ and the extracted LR properties are aggregated to obtain the final super-resolved result ($\mathbf{I}_{\mathrm{SR}}$).
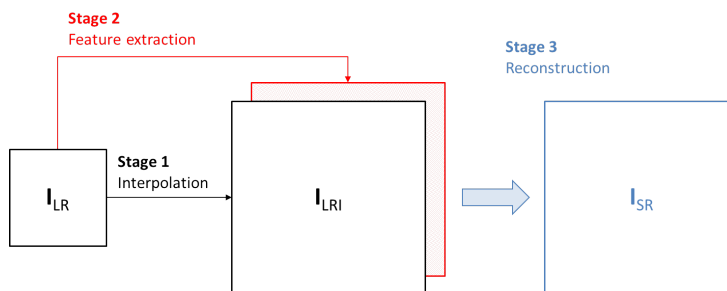

Figure 2.: Super-Resolution based on image reconstruction.

These methods upscale the LR image by a regular interpolation and then they

do image reconstruction to provide the same level of details we can observe in the input image. All the existing RE algorithms can be described in terms of the three aforementioned stages, however each particular method proposes the use of a particular set of tools to reconstruct the result. Specifically, we can highlight three kinds of techniques as some of the most relevant ones inside the RE category: (i) Iterative Back Projection, (ii) PSF Deconvolution and (iii) Gradient Profile. The followings sections present a description of those techniques according to the scheme proposed in Fig. 2.

3.1.1. Iterative Back Projection

The Iterative Back Projection (IBP) technique was first introduced in (Irani and S. 1991) as a general multi-frame super-resolution algorithm being extended to the single-frame approach afterwards becoming one of the most popular SR methods in the literature. The idea behind single-image IBP (we just focus on the single-image approach due to the scope of the work) is based on iteratively refining an initial guess of the super-resolved image. Throughout this iterative process, the reconstruction error between the LR image and the low-resolution version of the super-resolved image is minimised in order to obtain the final result. Eq. (2) provides the reconstruction error at a specific iteration ($\mathcal{E}_n$) given the initial low-resolution image ($\mathbf{I}_{\mathrm{LR}}$) and the super-resolved result of the previous iteration ($\mathbf{I}_{\mathrm{SR}}^{(n-1)}$).

$$\mathcal{E}_n = \mathbf{I}_{\mathrm{LR}} - \mathcal{D}(\mathcal{B}(\mathbf{I}_{\mathrm{SR}}^{(n-1)})) \tag{2}$$

where $\mathcal{D}$ and $\mathcal{B}$ are the decimation and blurring operators, respectively, as in Eq. (1). Three stages define the method: (i) initialisation, (ii) down-sampling and (iii) back-projection. In the initialisation stage (i), the super-resolved image is initialised with an interpolated version of the low-resolution input. After that, stages (ii) and (iii) are alternatively applied until a convergence condition is reached (e.g. a maximum number of iterations or a threshold in the reconstruction error). In the down-sampling stage (ii), the previous estimation of the super-resolved image $\mathbf{I}_{\mathrm{SR}}^{(n-1)}$ is blurred ($\mathcal{B}$) and decimated ($\mathcal{D}$) according to the corresponding imaging model to assess the reconstruction error $\mathcal{E}_n$ using Eq. (2). In the back-projection stage (iii), the current estimation of the super-resolved image $\mathbf{I}_{\mathrm{SR}}^{(n)}$ is obtained by projecting the reconstruction error to the previous estimation $\mathbf{I}_{\mathrm{SR}}^{(n-1)}$. Depending on the desired properties on the resulting image, a back-projection kernel can be applied over the projected error $\mathcal{E}_n$ to obtain smoother or sharper images.

3.1.2. PSF Deconvolution

The super-resolution methods based on the PSF deconvolution tackle the upscaling problem from a deblurring point of view. As it was introduced in Section 2, the imaging model introduce a blurring effect in the acquisition process caused by the operator $\mathcal{B}$. Specifically, this operator can be seen as the convolution of the sensor PSF with the HR scene. Using this approach, the super-resolution problem becomes one of undoing the blurring kernel effect.

It is theoretically possible to deconvolve a blurred image if we have access to the exact blurring kernel. However, the inherent noise that appears during any acquisition process together with the lack of information about the kernel, hinder that task. The instrument PSF may be unavailable or even it may not be completely

precise making the recovery of the exact solution impossible to be made.

Super-resolution methods based on PSF deconvolution can be defined as the concatenation of three phases: (i) interpolation, (ii) PSF estimation and (iii) deconvolution. The objective of the first stage (i) is to interpolate the input image to the same resolution as that of the output result. In the second stage (ii), the PSF is estimated from the input image $\mathbf{I}_{LR}$ in order to infer the imaging model blurring mask. Note that this phase can be omitted depending on the availability of the PSF. Finally, the third stage is in charge of removing the interpolated image blur according to the PSF in order to obtain the final super-resolved output $\mathbf{I}_{SR}$.

The deconvolution process is usually modelled by the following equation,

$$\mathbf{I}^* = \mathbf{I} * (\text{PSF}) + \mathcal{N} \tag{3}$$

where $\mathbf{I}^*$ and $\mathbf{I}$ represent the convolved and deconvolved images respectively, $*$ is the convolution operator, (PSF) represents the point spread function and $\mathcal{N}$ the additive noise. This deblurring problem has the same ill-posed nature as the SR problem. Therefore, many of the techniques used in SR are suitable to tackle the deconvolution problem as well. Non-blind approaches assume a specific model for the PSF, such us a normal distribution or a low-pass filter in the frequency domain, however blind approaches (Michaeli and M. 2013) are able to estimate the PSF from the convolved image by assuming some prior knowledge.

3.1.3. Gradient Profile

Gradient Profile (GP) SR methods take advantage of the fact that the gradient profile shape tends to be invariant across scales. Briefly speaking, GP describes a 1-D profile containing the gradient magnitude for each edge pixel along the gradient direction. Specifically, the general SR algorithm based on GP can be divided into three steps: (i) interpolation, (ii) GP extraction and (iii) reconstruction using a GP prior. In the first step (i), a regular interpolation algorithm is used to upscale the LR input image $\mathbf{I}_{LR}$. The second step (ii) is in charge of calculating the GP of the low-resolution image ($\nabla\mathbf{I}_{LR}$) as well as the interpolated image ($\nabla\mathbf{I}_{LRI}$) in order to obtain their sharpness. In the last step (iii), the GP of $\mathbf{I}_{LRI}$ is transformed according to the sharpness of $\nabla\mathbf{I}_{LR}$ and this transformed GP ($\nabla\mathbf{I}^*_{LRI}$) is used to reconstruct the super-resolved image. Eq. (4) shows the expression used to minimise the energy in both image domain (E_i) and gradient domain (E_g) with a control parameter β.

$$E(\mathbf{I}_{SR}|\mathbf{I}_{LR}, \nabla\mathbf{I}^*_{LRI}) = E_i(\mathbf{I}_{SR}|\mathbf{I}_{LR}) + \beta E_g(\nabla\mathbf{I}_{SR}|\nabla\mathbf{I}^*_{LRI}) \tag{4}$$

3.2. Super-Resolution based on Image Learning

Super-resolution based on image LEarning (LE) is probably one the most relevant research directions within the super-resolution field nowadays. The idea behind LE methods is to learn the potential relationships between low-resolution and high-resolution domains from an external training set and then to generate the final super-resolved image using this a priori knowledge.

Figure 3 shows the three stages the general LE scheme can be divided into. The first stage (learning) is aimed at learning the low-resolution to high-resolution pair relationships from a specific training set. In the second stage (estimation),

the LR input image is used to estimate those HR components that are related to the low-resolution elements appearing in $\mathbf{I}_{\mathrm{LR}}$. Eventually, the third stage (image generation) combines the estimated HR components of the input image to generate the final super-resolved image. For the sake of clarity, we use the term *generation* for LE methods and the term *reconstruction* for RE methods.
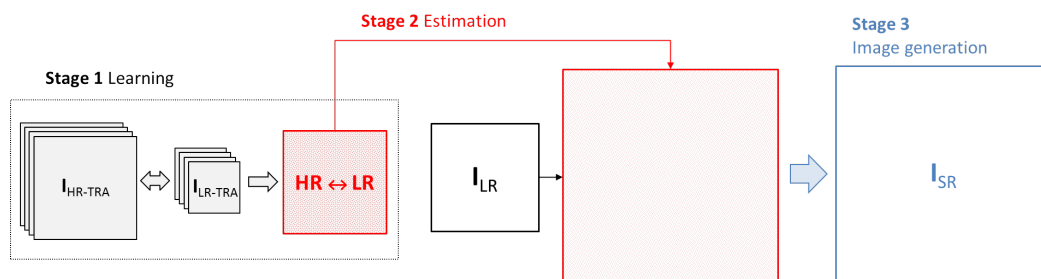
Figure 3.: Super-Resolution based on image learning.

The learnt relationships between HR and LR training details set the quality of the super-resolved images. Therefore, training data needs to be relevant enough to learn useful structures for the kind of input LR image ($\mathbf{I}_{\mathrm{LR}}$). Authors usually refer to the structure containing the relationship between low-resolution and high-resolution information as *dictionary* although it does not necessarily have to be implemented using an associative array. Somehow, it has to relate low-resolution and high-resolution components using a specific image characterisation. The most standard characterisation approach is the patch-based representation where pairs of HR/LR training images are cropped in order to learn functions to link them.

Each LE method uses its own strategies to learn the dictionary as well as to generate the final super-resolved output. Among the most popular techniques, the following three kinds of learning methods can be considered: (i) sparse coding, (ii) neighbourhood embedding and (iii) mapping.

3.2.1. Sparse Coding

Sparse Coding (SC) is a powerful image modelling technique which has been successfully used in several image restoration applications, being SR one of the most active domains. This kind of methods take advantage of the fact that natural images tend to be intrinsically sparse when they are characterised as a linear combination of small patches. In this way, SC algorithms pursue to obtain the sparse representation corresponding to the low-resolution input image and to generate the high-resolution output by means of the replacement of LR patches by their corresponding HR counterparts. In SC algorithms, we can highlight the following steps: (i) dictionary training, (ii) LR sparse code computation and (iii) SR output generation.

The first step is based on learning dictionaries for high-resolution ($\mathbf{D}_{\mathrm{h}}$) and low-resolution ($\mathbf{D}_{\mathrm{l}}$) image patches. These dictionaries are learnt by forcing the high-resolution training images ($\mathbf{I}_{\mathrm{HR^{TRA}}}$) and their low-resolution counterparts ($\mathbf{I}_{\mathrm{LR^{TRA}}}$) to share the same sparse codes. Equation (5) shows the formulation to train both dictionaries $\mathbf{D}_{\mathrm{h}}$ and $\mathbf{D}_{\mathrm{l}}$.

$$\arg\min_{\mathbf{D}_{\mathrm{h}},\mathbf{D}_{\mathrm{l}},\boldsymbol{z}} \frac{1}{h}||\mathbf{I}_{\mathrm{HR^{TRA}}} - \mathbf{D}_{\mathrm{h}}\,\boldsymbol{z}||_2 + \frac{1}{l}||\mathbf{I}_{\mathrm{LR^{TRA}}} - \mathbf{D}_{\mathrm{l}}\,\boldsymbol{z}||_2 + \lambda\left(\frac{1}{h} + \frac{1}{l}\right)||\boldsymbol{z}||_1 \quad (5)$$

In Eq. (5), $\boldsymbol{z}$ represents the sparse codes shared by both dictionaries, h and l are the dimensions of high-resolution and low-resolution patches respectively and λ represents the Lagrange multiplier to balance the sparsity of the solution. Note that in the training step the sparse codes $\boldsymbol{z}$ are not useful by themselves because they are only used to obtain the dictionaries. The first two terms in Eq. (5) correspond to the optimisation fitting part and the last term acts as a regulariser. Regarding the norms appearing in Eq. (5), the ℓ^2-norm is typically used for the fitting terms and the ℓ^1-norm, for the regularisation term but other norms may be used depending on the particular SC method.

The second step in SC algorithms is related to the computation of the low-resolution input image sparse codes. The solution of Eq. (6) is the $\boldsymbol{\alpha}$ coefficients from the trained dictionaries $\mathbf{D}_h$ and $\mathbf{D}_l$.

$$\arg\min_{\boldsymbol{\alpha}} \ \left|\left| \begin{bmatrix} F \ \mathbf{D}_l \\ \beta \ P \ \mathbf{D}_l \end{bmatrix} \boldsymbol{\alpha} - \begin{bmatrix} F \ \mathbf{I}_{\mathrm{LR}} \\ \beta \ \mathbf{w} \end{bmatrix} \right|\right|_2 + \lambda ||\boldsymbol{\alpha}||_1 \tag{6}$$

In this expression, F represents the feature extraction operator typically implemented by a high-pass filter, P is in charge of extracting the overlapping area between the target patch and the previously reconstructed image, β controls de compatibility of neighbouring patches and the fidelity to the initial low resolution image and $\mathbf{w}$ contains the values of the previously reconstructed high-resolution image on the corresponding overlapping area.

Finally, the high-resolution output is generated in the third step by using the $\boldsymbol{\alpha}$ sparse codes of the input low-resolution image over the high-resolution dictionary $\mathbf{D}_h$. Eq. (7) shows the expression to obtain this result.

$$\mathbf{I}_{\mathrm{SR}} = \mathbf{D}_h \ \boldsymbol{\alpha} \tag{7}$$

Even though these three steps are the standard process of SC algorithms, an extra post-processing stage is usually applied to enforce a global reconstruction constraint. Due to the ill-posed nature of the problem, the generated solution $\mathbf{I}_{\mathrm{SR}}$ is an approximation of the real high-resolution image $\mathbf{I}_{\mathrm{HR}}$. Therefore, the decimated version of the solution may differ from the LR input image. Eq. (8) allows eliminating this discrepancy by projecting the input image $\mathbf{I}_{\mathrm{LR}}$ onto the solution space.

$$\mathbf{I}_{\mathrm{SR}}^* = \arg\min_{\mathbf{I}_{\mathrm{SR}}} \ ||\mathcal{D} \ \mathcal{B} \ \mathbf{I}_{\mathrm{SR}} - \mathbf{I}_{\mathrm{LR}}||_2 + c||\mathbf{I}_{\mathrm{SR}} - \mathbf{I}_{\mathbf{SR_o}}||_2 \tag{8}$$

$\mathcal{D}$ and $\mathcal{B}$ are the decimating and blurring operators respectively, $\mathbf{I}_{\mathbf{SR_o}}$ represents the initial guess of the solution provided by Eq. (7) and $\mathbf{I}_{\mathrm{SR}}^*$ is the final output solution after the optimisation process which can be performed using a gradient descent approach. In a sense, the main aim of this global optimisation is to balance the fitting of the solution with the initial LR image and the solution with the initial high-resolution estimation by a factor c.

3.2.2.   Neighbourhood Embedding

Neighbourhood Embedding (NE) methods are inspired by recent manifold learning approaches which makes them also known as manifold-based methods. NE techniques assume that small image patches from low-resolution images and their high-resolution counterparts form low-dimensional non-linear manifolds with similar local geometries. This means that as long as enough samples are available, patches in the HR feature domain can be reconstructed as a weighted average of local neighbours using the same weights as in the LR feature domain. As in the case of any learning-based super-resolution algorithm, NE approaches can be described in terms of three stages: (i) manifold learning, (ii) LR neighbourhood reconstruction and (iii) SR output generation.

In the first step (manifold learning), both HR and LR manifolds are generated from a specific training set by means of minimising their corresponding reconstruction errors. Initially, a neighbourhood is found for each patch and then the reconstruction weights are computed using its own neighbourhood. The Euclidean distance is typically used to define the neighbourhood, however any other distance function may be used. Eq. (9) shows the reconstruction error for a patch $\mathbf{p}_i$,

$$\mathcal{E}_i = ||\mathbf{p}_i - \sum_{\mathbf{p}_j \in \mathbf{N_i}} \boldsymbol{x}_{i,j}\ \mathbf{p}_j||_2 \tag{9}$$

where $\mathbf{N_i}$ represents the $\mathbf{p}_i$ patch neighbourhood, $\mathbf{p}_j$ is an element of this neighbourhood and $\boldsymbol{x}_{i,j}$ is the array of reconstruction weights. The following constraints: $\sum \boldsymbol{x}_{i,j} = 1$ and $\boldsymbol{x}_{i,j} = 0$ for any $\mathbf{p}_j \notin \mathbf{N_i}$ apply to $\boldsymbol{x}_{i,j}$.

The second step (LR neighbourhood reconstruction) is based on using the same idea but just on the LR input image. That is, this step is aimed at finding the LR input image reconstruction weights in order to represent each of its patches as a combination of neighbours in the LR manifold by minimising Eq. (9).

Finally, the high resolution embedding is computed in the third step (SR output generation) using the low-resolution reconstruction weights over the high-resolution patches, preserving the local geometry of the reconstruction weights. In other words, the reconstruction weights $\boldsymbol{x}_{i,j}$ of the LR input image are used to seek patches $\mathbf{h}_i$ in the HR manifold. Eq. (9) shows the expression to achieve that goal.

$$\mathbf{h}_i = \sum_{\mathbf{h}_j \in \mathbf{N_i}} \boldsymbol{x}_{i,j}\ \mathbf{h}_j \tag{10}$$

3.2.3.   Mapping

Mapping (MP) based methods (also known as regression-based methods) consider super-resolution as a regression problem between high-resolution and low-resolution spaces. The underlying idea is to learn a mapping function from the input low-resolution images to the target high-resolution images based on a specific training set.Then, this learnt function can be used to generate the final super-resolved result from the low-resolution input image. In particular, MP methods are made up by three steps: (i) mapping function training, (ii) projecting the LR input image $\mathbf{I}_{\text{LR}}$ into the HR space and (iii) generating the final SR output.

In the literature, we can find all kind of techniques to perform the aforementioned regression. Bayesian networks, kernel-based methods or even neural networks are

some of the approaches which have been widely used to estimate the projection between LR and HR images. Even though many of these approaches use different theoretical tools to achieve this goal, the bottom line in the way the super-resolution process is performed is mainly the same. Because of that, we have decided to include all these methods in the same group despite the fact that other works may present a different classification. Let us review in this Section some of the most important mapping functions.

One of the first SR estimators appearing in the literature was the Nearest-Neighbour (NN) function which works as follows. In the training stage, pairs of LR and HR patches are collected. The input low-resolution image patches are compared against the stored collection in order to select the corresponding high-resolution version of the nearest ones. Finally, the super-resolved output is generated guaranteeing a certain spatial compatibility over the chosen HR patches.

Note the differences between a neighbourhood embedding approach (Section 3.2.2) and a SR method based on a NN mapping function. In the first case, there is a reconstruction process of input low-resolution patches using a specific neighbourhood and then these reconstruction codes are used in the high-resolution domain to generate the final super-resolved image. However, a NN-based mapping approach does not have this reconstruction. The rationale behind MP methods lies on having a function to directly connect low-resolution to high-resolution patches, and this function may be based on an NN approach. Depending on the embedding reconstruction process, it is possible to obtain the same result obtained by an NN mapping function but both approaches start from different assumptions.

Although NN-based methods are able to provide a reasonable good performance, these estimators tend to suffer from over-fitting in SR due to the inherent problem complexity. In general, one of the most effective ways to avoid over-fitting in machine learning is by regularisation. In a sense, regularising the estimator allow us to discard those solutions which are not general enough to generate a suitable output.

(Kim and Y. 2010) present a SR framework based on learning a mapping function via Kernel Ridge Regression (KRR) which uses the ℓ^2-norm in both the fitting and the regularisation terms. Despite the effectiveness of this approach, KRR is one of the simplest mapping functions and more complex schemes have been developed (He and W.-C. 2011).

More recent works, such us (Yang and M.-H. 2013; He and W.-C. 2011) train several simple mapping functions instead of using a single complex one. Eventually, SR is performed by applying the function associated to its closest cluster to each patch. Other works use a different kind of non-linear mapping. For instance, (Dong et al. 2015; Liu et al. 2016) use deep convolutional neural networks as mapping functions.

3.3. *Hybrid Super-Resolution*

Hybrid (HY) SR algorithms have features from both RE and LE methods in order to address some of their shortcomings. In the case of RE methods, the unavoidable blur generated by the initial interpolation together with the lack of important high-frequency details in the low-resolution input image limits their effectiveness to small magnification factors (Baker and T. 2002). In the case of LE methods, the constraint is generated by the availability of a suitable training set. Even though image learning-based methods may be able to learn spatial details impossible to recover by a reconstruction approach, LE methods usually require a significant

amount of relevant high-resolution images and this last requirement is not always possible. Hybrid methods try to reach an agreement between avoiding the use of external HR training images and performing a training process to learn LR/HR relations.

All hybrid SR methods somehow exploit the redundancy property pervading natural images. This property refers to the fact that natural high-resolution images tend to contain repetitive structures within the same scale and over different scales as well. In general, HY algorithms can be summarised in the following three stages: (i) patch redundancy search, (ii) HR patch extraction and (iii) SR output generation. The first step (patch redundancy search) is aimed at both generating several lower scale images and extracting those patches which tend to appear across scales. Each particular HY approach defines its own assumptions about the imaging model to generate the lower scale images and about the criteria to seek repetitive patches (Glasner, S., and M. 2009; Freedman and R. 2011).

4. Super-resolution quality assessment

The effectiveness of super-resolution algorithms relies on the quality of the super-resolved output images. Nevertheless, how to quantify the super-resolution result remains an open issue. It was widely accepted in the past that the best way to evaluate image quality was by the user's opinion. However, this kind of evaluation may be biased, slow and expensive, being often impractical in remote sensing. Besides, many applications are not user-aimed, i.e. images are not used by users but by programs in order to perform a specific task, for instance soil classification from remote sensing imagery. As a result, over the past decades an important effort has been made to develop automatic tools to perform such evaluation. Even though there are several image quality metrics available in the literature, each one tends to focus on a particular kind of visual features and therefore some metrics may become more useful than others depending on the final application of the super-resolved images.

In general, two kinds of quality metrics can be considered: (i) reference-based metrics and (ii) no-reference-based metrics. The former group uses ground-reference images to compare the super-resolved output against its *true* high-resolution image SR algorithms aim at recovering. On the other hand, the latter group is able to assess the super-resolved images without the use of any reference image. In the following sections, we are going to review some of the most popular metrics of both groups.

4.1. *Image quality metrics with reference*

Image quality metrics with reference compare the properties of a given image $\mathbf{I}$ with respect to its distortion-free version $\mathbf{R}$ by measuring the deviation between both images. In particular, the computation of these deviations can be focused on two different aspects, namely spatial and spectral distortions. Whereas spatial measures are more concerned about the similarity between geometric structures of both images, the spectral *indexes* try to detect radiometric differences.

One of the most well-known spectral indexes is the Spectral Angle Mapper (SAM) (Yuhas, H., and W. 1992), which is widely used in pan-sharpening spectral assessment (Vivone et al. 2015). Specifically, SAM considers each spectral band as a coordinate axis and then it computes the average angle between the pixels of $\mathbf{I}$ and

R images. Eq. (11) shows the expression defining this index.

$$\text{SAM}(\mathbf{I}, \mathbf{R}) = \frac{1}{N} \sum_i^N \arccos \frac{\boldsymbol{I}_i \cdot \boldsymbol{R}_i}{||\boldsymbol{I}_i|| \ ||\boldsymbol{R}_i||} \quad \in [0, \pi] \tag{11}$$

where N is the total number of pixels in each image. Note that the ideal value of SAM is 0. Other indexes account for the spectral and spatial distortion simultaneous measurement. A classical error measure which has been widely used in image quality assessment is the Root Mean Square Error (RMSE), defined as follows,

$$\text{RMSE}(\mathbf{I}, \mathbf{R}) = \sqrt{\frac{1}{K \cdot N} \sum_j^K \sum_i^N (I_i^j - R_i^j)^2} \quad \in [0, \infty) \tag{12}$$

where the index j is used to identify each one of the K image bands. However, the main drawback of this index is based on the fact that it is very sensitive to outlier values and other measures are usually preferred. Another widely used index is the Peak Signal-to-Noise Ratio (PNSR) which is computed according to Eq. (13).

$$\text{PSNR}(\mathbf{I}, \mathbf{R}) = 20 \ \log_{10} \frac{255}{(\text{RMSE})} \quad \in (0, \infty) \tag{13}$$

A higher PSNR value represents a better image quality and its domain is all positive real numbers greater than 0. Even though PSNR seems to be the standard index in general purpose SR, it does not take into account the scaling factor to evaluate the super-resolved image. However, other measures defined within the Pansharpening context try to take advantage of this information. This is the case of the Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) measure (Veganzones et al. 2016), defined in Eq. (14).

$$\text{ERGAS}(\mathbf{I}, \mathbf{R}) = \frac{100}{S} \sqrt{\frac{1}{K} \sum_j^K \left(\frac{\text{RMSE}(\mathbf{I}^j, \mathbf{R}^j)}{\overline{\mathbf{I}^j}} \right)^2} \quad \in [0, \infty) \tag{14}$$

In this expression, S is the SR scaling factor and the over-line operator is used to represent the mean value of the band I^j. The best ERGAS index value is 0 and an increasing value means more distortion for I. Although RMSE-based measures are appealing because of their simplicity, they are not able to capture the characteristics of the Human Visual System (HVS) and hence other quality measures have been developed. In particular, the Correlation Coefficient (CC) is one of them. Eq. (15) shows how it is computed.

$$\text{CC}(\mathbf{I}, \mathbf{R}) = \frac{\sigma_{\text{IR}}}{\sigma_{\text{I}} \ \sigma_{\text{R}}} = \frac{\sum_{i,j}^{N,K} (I_i^j - \overline{\mathbf{I}^j})(R_i^j - \overline{\mathbf{R}^j})}{\sqrt{\sum_{i,j}^{N,K} (I_i^j - \overline{\mathbf{I}^j}) \ \sum_{i,j}^{N,K} (R_i^j - \overline{\mathbf{R}^j})}} \quad \in [-1, 1] \tag{15}$$

The CC index can be assessed computing the covariance matrix between both

14

images and dividing it by their standard deviation. In a sense, this index measures the linear relationship between two images normalising their range values. The best CC value is 1, implying that both images are linearly correlated. In remote sensing, it is common to use a particular case of CC which only takes edge information into account. That is, the spatial Correlation Coefficient (sCC) computes the CC index over the edges detected (Canny, Sobel or any other edge detector can be used to that purpose) in both $\mathbf{I}$ and $\mathbf{R}$ images. Another popular metric is the so-called Q-index (universal image Quality index) computed as follows.

$$\mathrm{Q}(\mathbf{I}, \mathbf{R}) = \sum_j^K \left(\overbrace{\frac{\sigma_{\mathrm{IR}}}{\sigma_{\mathrm{I}} \ \sigma_{\mathrm{R}}}}^{a} \overbrace{\frac{2 \ \bar{\mathbf{I}} \ \bar{\mathbf{R}}}{(\bar{\mathbf{I}})^2 \ (\bar{\mathbf{R}})^2}}^{b} \overbrace{\frac{2 \ \sigma_{\mathrm{I}} \ \sigma_{\mathrm{R}}}{(\sigma_{\mathrm{I}})^2 \ (\sigma_{\mathrm{R}})^2}}^{c} \right)_j \quad \in \ [-1, 1] \qquad (16)$$

This index gathers three different properties in the image evaluation: (a) correlation, (b) luminance and (c) contrast. The first term (a) corresponds to the Correlation Coefficient expression and therefore it is in charge of measuring the linear relation between both $\mathbf{I}$ and $\mathbf{R}$ images. The second term (b) and the third one (c) aim at comparing luminance and contrast, respectively. Note that in the case of multi-spectral images, the Q-index can be computed as the average value among spectral bands. An extension of the Q-Index was proposed in (Wang et al. 2004) in order to avoid instability around null values. Eq. (17) shows the expression associated to the Structural SIMilarity (SSIM) metric.

$$\mathrm{SSIM}(\mathbf{I}, \mathbf{R}) = \sum_j^K \left(\frac{(2 \ \bar{\mathbf{I}} \ \bar{\mathbf{R}} + C_1)(2 \ \sigma_{\mathrm{IR}} + C_2)}{(\bar{\mathbf{I}}^2 + \bar{\mathbf{R}}^2 + C_1)(\sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{R}}^2 + C_2)} \right)_j \quad \in \ [-1, 1] \qquad (17)$$

According to (Wang et al. 2004), C_1 and C_2 constants should be set to $(K_1 L)$ and $(K_2 L)$ respectively, where K_1 and K_2 are values close to 0 and L represents the image dynamic range. Similarly to the Q-index, the SSIM index range of values is $[-1, 1]$ and it can be computed as the average over bands in the case of multi-spectral images.

Even though the availability of ground truth images is the ideal scenario to perform the image quality assessment, this is not possible in many real-life applications. However, alternative protocols have been developed to enable the use of metrics with reference image when ground truth images are not available.

The reduced-resolution protocol (Wald, T., and M. 1997) was initially defined within the Pansharpening context but it can be easily extended to SR as well. In particular, this protocol tests two properties: (a) consistency and (b) synthesis (see Fig. 4). On the one hand, the consistency property is based on the fact that the low-resolution version of the super-resolved image has to be as similar as possible to the initial input image. In other words, the LR input image is the ground truth of the super-resolved image after using the imaging model operators. On the other hand, the synthesis property performs the evaluation at a lower scale. That is, the initial LR image is reduced following the imaging model to generate a smaller low-resolution image. Then, this smaller image is used to perform SR to the initial LR scale having its corresponding reference image available.

(a) Consistency

(b) Synthesis
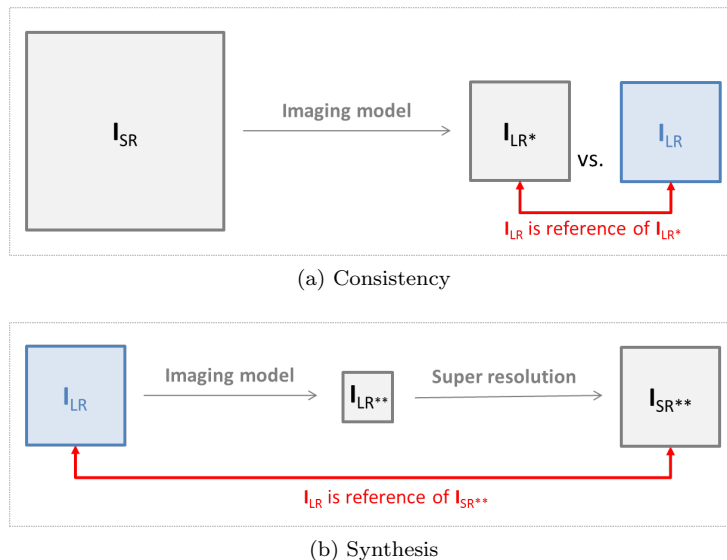
Figure 4.: Reduced-Reference assessing protocol.

4.2. *No-reference image quality metrics*

The availability of ground truth images provides the ideal scenario to perform image quality assessment, however several no-reference metrics have also been proposed to automatically predict the image quality without any kind of ground truth image. Broadly speaking, it is possible to summarise the existing no-reference metrics in two different families (Nouri et al. 2013): (1) distortion-based and (2) training-based metrics. On the one hand, the metrics belonging to the first group use a specific distortion model to quantify image peculiarities, such us blocking artifacts, ringing or blur. On the other hand, training-based metrics require an initial training process to learn high-quality image statistics and, using this information, they are able to quantify the image quality.

Over the past years, several no-reference metrics belonging to both families have been presented. However, training-based metrics have shown to provide results that are more correlated to human judgements than distortion-based ones (A.C. 2010).

One of the most popular training-based metrics is Blind Referenceless Image Spatial QUality Evaluator (BRISQUE) (Mittal, K., and C. 2012). This measure is based on the observation that normalised luminance values tend to follow a unit Gaussian distribution for distortion-free natural images. Specifically, BRISQUE computes the normalized luminance values by local mean subtraction and contrast normalisation, fitting these values to a Generalised Gaussian Distribution (GGD) afterwards. This learnt GGD is finally used to predict the type of distortion as well as the perceptual quality. The implementation provided by the authors uses a Support Vector Regressor (SVR) to learn the quality scores from a given training set but any other regressor could be used instead. Another similar popular training-based metric is the Natural Image Quality Evaluator (NIQE). In this case, the NIQE index uses a distortion-free training set to fit a multivariate Gaussian (MVG) distribution. Then, image quality is computed as the distance between the input image features and the MVG distribution previously learnt. Regarding the kind of features, the NIQE metric uses features that are similar to those used by BRISQUE, that is, normalised luminance values fitted to Gaussian distributions. The main difference between BRISQUE and NIQE approaches lies on the fact that the second one does not require any regressor calibration.

16

5. Experiments

This section describes the experiments which have been carried out in order to validate and compare SR algorithms performance in challenging remote sensing scenarios. Specifically, the image databases used for the experiments are introduced in subsection 5.1. The SR methods are enumerated in subsection 5.2. Finally, the experimental protocol and the obtained results are shown in subsections 5.3 and 5.4, respectively. Datasets and SR codes used in this work are publicly available in *http://www.vision.uji.es/srtoolbox/*.

5.1. *Datasets*

As outlined in the introduction, the main objective of this work is to provide a practical overview about single-frame super-resolution algorithms performance in typical remote sensing imagery. Hence, the perfect scenario to highlight practical differences among methods is by super-resolving images of the same area captured using different sensors resolution. Besides, the availability of ground truth images is also important to enable a full-reference evaluation in order to quantify the deviation between super-resolved images and the reference high-resolution ones.

Nevertheless, an important *volume* of the total amount of remote sensing images currently acquired belongs or is available only through private companies rights, which hinders the possibility to find public domain images containing the same area and, at the same time, acquired by different sensors. Moreover, ground truth images are often unavailable in the case of satellite images because of the own sensor limitations. In fact, the lack of HR images is the main motivation behind the use of SR methods in remote sensing.

Taking these reasons into account, we have simulated different databases from a specific real set of high-resolution aerial images. Note that the difference among sensors lies on their imaging models. Therefore, it is possible to generate several databases from the same set of ground truth images by applying different imaging models. This approach gives us three main advantages. First of all, ground-reference images are available which allows us to apply both full-reference and no-reference evaluation protocols. Second, we can easily generate as many databases as we need by using different imaging models, for instance changing the noise level or the down-sampling scale. Thirdly, imaging model features are controlled, which allows us to analyse SR performance under acquisition model changes.

The reference images used in this work have been selected from the open-access orthoimages of the National Aerial Ortophoto Program (PNOA) of the Spanish *Ministerio de Fomento* (Arozarena, G., and N. 2005). These RGB images have a resolution of 0.25 mpp (meters per pixel), higher than those of satellites like WorldView-3 (PAN: 0.31 mpp, MS: 1.24 mpp) or GeoEye-2 (PAN: 0.34 mpp, MS: 1.36 mpp), and they are publicly available from the National Geographic Institute (IGN) website (*http://pnoa.ign.es/*).

Specifically, we considered a total of 20 high-resolution 512×512 images from Alicante (Valencian Community, Spain) belonging to four different area types: (i) agricultural, (ii) industrial, (iii) urban and (iv) forest. Figure 5 shows the reference images considered in the experiments.

From these HR images, we generated six different LR datasets by changing the imaging model parameters used in each case. Datasets are described below.

- **01LR2xbd**: In this case, low-resolution images have been obtained by using

17
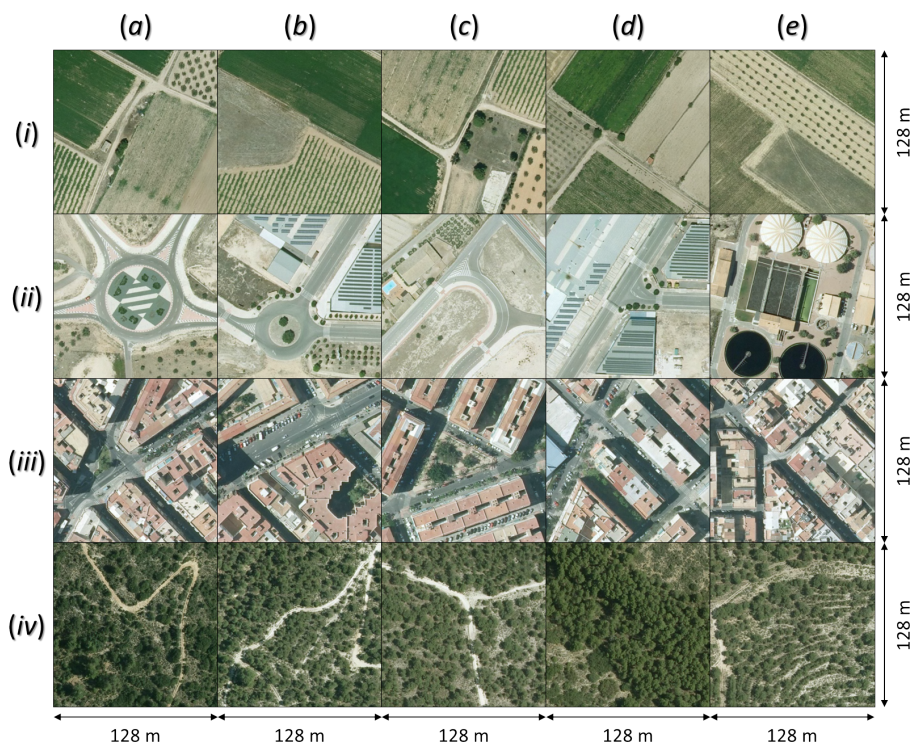
Figure 5.: High-Resolution reference images (RGB, 512×512, 0.25mpp). Five different image folds: from (a) to (e), of four different areas: from (i) to (iv).

a Gaussian PSF ($\mu = 0, \sigma = 1$) as a blurring operator ($\mathcal{B}$), a 2x size reduction in the decimation operator ($\mathcal{D}$), with a zero noise level ($\mathcal{N} = 0$). Note that the resulting LR images are 256×256 and their resolution is 0.5mpp.

- **02LR4xbd**: This database has been obtained using a similar configuration as the previous one but using a 4x decimation. In particular, we have consecutively applied it (twice) the imaging model described in '01LR2xbd' to obtain LR images with a size of 128×128 and a resolution of 1mpp.

- **03LR2xbdn01**: The generation of this database follows the same process described in '01LR2xbd' but in this case we have used an additive white Gaussian noise with $\sigma = 0.01$. The final images have a size of 256×256 and a resolution of 0.5mpp

- **04LR4xbdn01**: For this database, we have followed the process described in '02LR4xbd' adding a $\sigma = 0.01$ white Gaussian noise to the final images. In this case, the images of this database have a size of 128×128 and a resolution of 1mpp.

- **05LR2xbdn05**: This database has been generated following the procedure described in '03LR2xbdn01' but using a $\sigma = 0.05$ additive white Gaussian noise. The images belonging to this database are 256×256 and 0.5mpp.

- **06LR4xbdn05**: The imaging model used for this database is the same that one of '04LR4xbdn01' but considering a $\sigma = 0.05$ additive white Gaussian noise, resulting 128×128 images with 1mpp.

It should be noted that the parameters to generate these databases have been chosen following the most common scenarios in remote sensing imagery.

5.2. *Super-Resolution Toolbox*

This section provides a brief description of the SR methods considered in the experimental part of the work. Specifically, a total of 27 single-frame super-resolution methods (see Table 1) have been tested over the six aforementioned databases. Besides, Lanczos-3 interpolation algorithm has been used as an up-scaling baseline. All these methods have been selected for this comparative work mainly because they are among the most popular single-frame SR methods in the literature and besides their implementations are publicly available either on the Internet or from the authors. Subsequently, we highlight the main features of the methods in terms of the information provided in Section 3.

Table 1.: Summary table that captures the SR approaches tested in this work.

Single-image super-resolution families		Tested methods	#
Image reconstruction (RE)	**Iterative back projection (IBP)**	01IBP (Irani and S. (1991))	1
	PSF deconvolution	02FIU (Shan et al. (2008))	4
		23DLU (Lucy (1974))	
		24DRE (Gonzalez and Woods (2006))	
		25FSR (Zhao et al. (2015))	
	Gradient profile (GP)	03GPP (Jian, Xu, and Shum (2008))	1
Image learning (LE)	**Sparse coding (SP)**	06VSR (Yang et al. (2010))	6
		08SDS (Dong et al. (2011))	
		13SIP (Yang et al. (2008))	
		14SIS (Zeyde, M., and M. (2012))	
		20MDL (Purkait and B. (2013))	
		22BSR (Polatkan et al. (2015))	
	Neibourhood embedding (NE)	10ANR (Timofte, V., and L. (2013))	5
		15GLR (Timofte, V., and L. (2013))	
		16NELS (Timofte, V., and L. (2013))	
		17NENNLS (Timofte, V., and L. (2013))	
		18NELLE (H., T., and Y. (2004))	
	Mapping (MP)	05SRP (Kim and Y. (2010))	6
		09DSF (Yang and M.-H. (2013))	
		11GPR (He and W.-C. (2011))	
		19JOR (Dai, R., and L. (2015))	
		21CNN (Dong et al. (2014))	
		27SCN (Liu et al. (2016))	
Hybrid methods (HY)		04SRSI (Glasner, S., and M. (2009))	4
		07LSE (Freedman and R. (2011))	
		12BDB (Michaeli and M. (2014))	
		26TSE (Huang, A., and N. (2015))	

- **00IntL3 (Lanczos-3 Interpolation) (Turkowski 1990)**: In general, Lanczos interpolation uses a normalised version of a windowed cardinal sine function. In the particular case of Lanczos-3, this *sinc* function is made of 5 lobes and it has shown to offer one of the best compromises in terms of reduction of aliasing, sharpness and minimal ringing (Turkowski 1990). Due to these reasons, we have selected Lanczos-3 interpolation as the baseline function for the experiments. Specifically, we use the implementation provided in Matlab R2013a Image Processing Toolbox.
- **01IBP (Iterative Back Projection) (Irani and S. 1991)**: This algorithm follows the approach presented in Section 3.1.1. We used the implementation provided in (Yang, C., and M.-H. 2014) which uses a 5×5 Gaussian kernel with $\sigma = 0.6$ to perform the initial image interpolation and a total number of 100 back-projection iterations.
- **02FIU (PSF Deconvolution) (Shan et al. 2008)**: This method integrates the scheme introduced in Section 3.1.2 with a feedback-control framework to

constrain the image up-sampling process. Initially, the LR input image is interpolated to the desired resolution by a bi-cubic interpolation. Then, this up-sampled image is deconvolved using the non-blind deconvolution method proposed in (Shan, J., and Agarwala 2008). The implementation we use for the experiments has been provided by (Shan et al. 2008). In particular, this code uses a total number of 10 iterations and the Gaussian PSF has the following parameters, $\sigma = 1.05$ for a scaling factor of 2x and $\sigma = 1.5$ for a scaling factor of 4x.

- **03GPP (Gradient Profile) (Jian, Xu, and Shum 2008)**: This algorithm exploits the idea shown in Section 3.1.3. The implementation provided by the authors has the following characteristics. First, it uses bi-cubic interpolation to perform the initial image up-sampling process. Second, each gradient profile distribution is fitted to a Generalized Gaussian Distribution (GGD) which adds an extra shape control parameter to the normal distribution. Third, the transformation of the gradient is performed by multiplying the gradient profile of the LR input image ($\nabla \mathbf{I}_{LR}$) by a ratio r which depends on the scaling factor (see (Jian, Xu, and Shum 2008) for more details).

- **04SRSI (Hybrid) (Glasner, S., and M. 2009)**: The idea behind this method is based on the hybrid approach presented in Section 3.3. Specifically, the code provided by the authors assumes a Gaussian PSF and it builds a 6-level resolution pyramid by blurring and down-sampling the initial LR image using a scaling factor of 1.25. Then, each patch of the input image is searched within the scales from higher to lower resolution using the Nearest-Neighbour algorithm.

- **05SRP (Mapping) (Kim and Y. 2010)**: This approach is a mapping-based method with the following features. In the training stage, a Kernel Ridge Regressor (KRR) is trained using a specific set of LR/HR patch pairs. The implementation used for the experiments has been extracted from the source code of (Yang, C., and M.-H. 2014).

- **06VSR (Sparse Coding) (Yang et al. 2010)**: This method corresponds to the procedure explained in Section 3.2.1. The code provided by (Yang, C., and M.-H. 2014) uses 1st and 2nd order derivatives as a feature extractor operator (F) to represent 5×5 patches. Besides, the patch overlapping is set to 4 pixels and the sparsity factor β to 0.2. Finally, the $\mathcal{B}$ operator in the post-processing step is considered as a 5×5 Gaussian blurring mask with $\sigma = 1$.

- **07LSE (Hybrid) (Freedman and R. 2011)**: This work is based on a similar approach to the one detailed in 04SRSI. That is, it exploits the self-similarity property but with a slight refinement to optimise the searching process. The implementation provided by the authors uses a custom non-dyadic filter bank as a blurring operator and the desired magnification factor is upper approximated by small scaling steps compatible with that filter, 5/4, 4/3 and 3/2. Finally, a bi-cubic decimation is performed to achieve the target resolution.

- **08SDS (Sparse Coding) (Dong et al. 2011)**: This work presents an extension of the procedure explained in 06VSR. In particular, it introduces two improvements: (i) an initial training data clustering process and (ii) an adaptive regularisation in the generation step. In the code provided by the authors the images are initially processed using a Gaussian high-pass filter, patches are 7×7 sized with 5 pixels of overlapping and the number of clusters as well as the number of Auto-Regressive functions per cluster is set to 200.

- **09DSF (Mapping) (Yang and M.-H. 2013)**: This mapping method is based on learning several simple functions, instead of a single complex one, to relate high-resolution and low-resolution structures. Eventually, the super-resolved image is obtained by using the mapping function associated to its closest cluster, on each input LR patch. The implementation provided by the authors considers a patch size of 7×7 and removes each patch mean in order to characterise them.

- **10ANR (Neighbourhood Embedding) (Timofte, V., and L. 2013)**: This technique is an extension of the neighbourhood embedding idea presented in Section 3.2.2. In fact, it aims at fusing both Neighbourhood Embedding (NE) and Sparse Code (SC) approaches. The code provided by the authors uses a patch characterisation based on 1st and 2nd order gradient and besides it applies Principal Component Analysis (PCA) to reduce the data dimensionality. The patch size is 3 with 1 pixel of overlapping.

- **11GPR (Mapping) (He and W.-C. 2011)**: Even though we have classified this method as a mapping one because it uses a Gaussian Process Regression (GPR) function to perform the SR task, it could also be considered a reconstruction or even an hybrid method because it extends the mapping approach in order to avoid the use of external training examples. The implementation used in the experiments has been provided by the authors of the method and it has the following features: a patch size of 20 with a $2/3$ overlapping factor and a Gaussian PSF with $\sigma = 0.6$.

- **12BDB (Hybrid) (Michaeli and M. 2014)**: This method is a de-blurring method that we have adapted to SR by carrying out an initial Lanczos-3 interpolation to the target resolution. We think it is important to test this technique because it follows a similar hybrid approach to the one proposed by (Michaeli and M. 2013) whose implementation is not available. In particular, this de-blurring method takes advantage of the cross-scale patch recurrence property presented in Section 3.3: patches in sharp natural images tend to co-occur across scales whereas patches in blurred images do not. The code of this de-blurring method has been provided by the authors. Specifically, it uses a 5×5 patch and a 13×13 kernel. The image scaling pyramid is constructed using a $3/4$ down-scaling factor and the *sinc* function until the kernel is smaller than the patch size.

- **13SIP (Sparse Coding) (Yang et al. 2008)**: This algorithm follows the sparse coding approach presented in Section 3.2.1 but simplifying the dictionary training step. In particular, this technique corresponds to an initial version of 06VSR. The main difference between 06VSR and 13SIP lies on the fact that each method learns the dictionaries in a different way. The implementation provided by the authors considers a patch size of 3 with 1 pixel of overlapping. The kind of extracted features, the sparsity factor and the blurring mask in the post-processing step are the same to those considered in 06VSR.

- **14SIS (Sparse Coding) (Zeyde, M., and M. 2012)**: This sparse coding method is an extension of the method explained in 06VSR. Specifically, the most important improvements are the use of PCA in the image characterisation process and the use of an optimised dictionary training step. The code provided by the authors have the same configuration to the one of 13SIP.

- **15GLR (Neighbourhood Embedding) (Timofte, V., and L. 2013)**: This approach is essentially the same to the one described in 10ANR but using different neighbourhood constraints. In particular, this variant uses a global

21

neighbourhood to represent each element of the dictionary as a combination of neighbours. The regularisation type and the method parameters are the same to those described for 10ANR.

- **16NELS (Neighbourhood Embedding) (Timofte, V., and L. 2013)**: This method is another variant of 10ANR. In this case, the regression, where each dictionary atom is represented as a linear combination of neighbours, is tackled as a Least-Squares (LS) problem without any kind of regularisation operator. The kind of neighbourhood and method parameters are those of 10ANR.

- **17NENNLS (Neighbourhood Embedding) (Timofte, V., and L. 2013)**: This NE method is a variant of 16NELS where the LS problem is constrained to positive values. The rest of the parameters are the same to those explained in 16NELS.

- **18NELLE (Neighbourhood Embedding) (H., T., and Y. 2004)**: This algorithm is based on the most classical NE approach introduced in Section 3.2.2. This method does not train any dictionary to perform the embedding, that is, the dictionary consists of training patches themselves. The rest of the parameters and settings of this method are the same as those of 10ANR.

- **19JOR (Mapping) (Dai, R., and L. 2015)**: The idea behind this mapping approach is similar to that one described in 09DSF. Specifically, this method is made up of two stages: (i) jointly-optimized regressor learning and (ii) adaptively HR generation. The implementation provided by the authors uses 1st and 2nd order derivatives as features, 3×3 patches with overlapping and 20 Expectation-Maximisation iterations to cluster the training data.

- **20MDL (Sparse Coding) (Purkait and B. 2013)**: This method presents an extension of the classical SC approach introduced in Section 3.2.1. The rationale behind this algorithm is similar to that of 19JOR but from a sparse coding point of view. The code used for the experiments was given by the authors. It uses a 7×7 patch encoded using a Bag-of-Words (BoW) approach over 1st and 2nd order derivatives. The number of dictionaries is set to 20 by default and the scaling factor is a constant value of 1.25 which has to be used several times to reach higher up-scaling ratios. Finally, the super-resolved output is post-processed considering a 3×3 Gaussian PSF with $\sigma = 0.2$.

- **21CNN (Mapping) (Dong et al. 2014)**: This SR technique is inspired by the mapping approach explained in Section 3.2.3. In this case, a Convolutional Neural Network (CNN) is used to relate low-resolution structures to high-resolution ones. In the training stage, the method up-scales the LR input images by a bi-cubic interpolation, and then a 3-layer CNN is trained to learn a end-to-end mapping between these interpolated versions and their ground-reference counterparts.

- **22BSR (Sparse Coding) (Polatkan et al. 2015)**: This approach defines a probabilistic graphical model to perform the super-resolution task. Specifically, this Bayesian SR method can be summarised in four main steps: (1) coupled dictionary training, where high-resolution and low-resolution dictionaries are learned, (2) sparse low-resolution factor estimation, (3) super-resolved image generation, and (4) a final post-processing step following Eq. (8). The code provided by the authors uses an 8×8 patch size and a Gaussian PSF for the post-processing step.

- **23DLU (PSF Deconvolution) (Lucy 1974)**: In this method, we use the Richardson-Lucy (RL) deconvolution algorithm presented in (Lucy 1974) to tackle the SR problem from a de-blurring point of view (Section 3.1.2). For

the experiments, we use Lanczos-3 interpolation (00IntL3) to up-scale the initial LR image $\mathbf{I}_{\text{LR}}$ to $\mathbf{I}_{\text{LRI}}$ and a Gaussian PSF with $\sigma = 1$ to run the RL deconvolution function available in Matlab R2013a.

- **24DRE (PSF Deconvolution) (Gonzalez and Woods 2006)**: This method follows a similar approach to 23DLU but using a different deconvolution algorithm. In this case, we use the regularised deconvolution function provided in Matlab R2013a. This deconvolution technique is a linear method which uses a least-squares solution with a Laplacian regularization to preserve image smoothness. This kind of regularisation aims at eliminating noise artifacts. The configuration used for the experimentation is the same than one described in 23DLU.

- **25FSR (PSF Deconvolution) (Zhao et al. 2015)**: This SR method is an extension of the classical deconvolution approach (Section 3.1.2) where the decimation ($\mathcal{D}$) and blurring ($\mathcal{B}$) operators in Eq. (1) (imaging model) are unified. In (Zhao et al. 2015), three different regularisation strategies are tested with the proposed technique: (a) ℓ^2-norm (Tikhonov regularization), (b) ℓ^2-norm (Total Variation) and (c) ℓ^1-norm in the Wavelet domain. We tested the code provided by the authors using the Total Variation regularisation (b) because it provides one of the best results in (Zhao et al. 2015).

- **26TSE (Hybrid) (Huang, A., and N. 2015)**: The idea of this method is essentially the same to that presented in Section 3.3 but extending the self-similarity assumption. The implementation provided by the authors uses a 5×5 patch size and 20 post-processing iterations with a Gaussian PSF ($\sigma = 1.2$). The number of levels to achieve a scaling ratio of 2x and 4x are 3 and 6, respectively.

- **27SCN (Mapping) (Liu et al. 2016)**: This technique is an extension of 21CNN in order to combine the strengths of deep neural network and sparse coding. In particular, this approach defines a Convolutional Neural Network (CNN) layer structure similar to that described in 21CNN for patch extraction and reconstruction. The key difference is based on the kind of unit used to learn the non-linear mapping between interpolated LR patches and their HR counterparts. The code available in the authors' website normalises patches by their mean and variance and uses a HR patch size of 5×5.

5.3. *Experimental Protocol*

This section provides the details about how images described in Section 5.1 have been super-resolved using the SR methods appearing in 5.2 and assessed using metrics in Section 4. For the experiments, a total of 6 simulated databases have been generated using 6 different imaging models (described in Section 5.1) over the ground-reference images showed in Fig. 5. Note that each database contains 20 high-resolution ground-reference images (512×512) and their low-resolution counterparts (256×256 or 128×128 depending on the database). Each database has been split into 5 folds, columns from (a) to (e) in Fig. 5, in order to follow a cross-validation protocol for the results. That is, low-resolution images of a specific fold have been super-resolved using the rest of the folds as training and besides image quality metrics have been averaged over folds. Logically, only those SR methods requiring a training stage (image learning methods Section 3.2) have been trained.

Regarding the scaling factor, the target resolution of each experiment has been set to reach the ground-reference resolution, that is, 512×512. Note that a scaling factor of 2x has been used in databases with 256×256 low-resolution images and a

factor of 4x in databases with 128×128 images. Finally, the quality of the super-resolved images $\mathbf{I}_{SR}$ has been evaluated using the 9 metrics presented in Section 4 with a full-reference assessment protocol, i.e. using the ground-reference images $\mathbf{I}_{HR}$ for those metrics with reference. In addition, the computational time of each super-resolution method has been reported to evaluate the efficiency of the methods Training time has not been taken into account because this process can be carried out off-line.

5.4. *Results*

For each one of the datasets described in Section 5.1: $\{\mathbf{01LR2xbd}, \ldots\}$, the mean value for each figure of merit was obtained considering a 5-fold cross-validation strategy (folds from (a) to (e) of Figure 5). Then, the final figure of merit was the mean over the 4 different types of images (types from (i) to (iv) of Figure 5). Two different comparisons were made afterwards:

(1) Effect of a change in the resolution factor on the reconstruction quality. We compared the super resolved image quality results obtained when a 2x resolution factor was applied compared to the use of a 4x resolution factor. In this comparison, no noise was added to the images.
(2) Effect of noise on reconstruction quality for a specific super-resolution factor.

Comparisons were made using three of the figures of merit proposed, i. e., SAM, RMSE and PSNR, as given by Eqs. 11, 12 and 13. As explained in Section 4.1, SAM is a measure of the spectral change between two images. On the other hand, RMSE and PSNR are two widely accepted measures when assessing image quality for the case when there is a reference image.

Tables 2 to 7 show the results of the best three SR methods for each type of image (from (i) to (iv)) for each one of the datasets described ($\{\mathbf{01LR2xbd}, \ldots\}$) and for all the figures of merit analysed in this paper. Baseline method **00IntL3 (Turkowski 1990)** is also included. **08SDS (Dong et al. 2011)** method is always amongst the best three methods in all the figures of merit that use a reference image for all the datasets but in **05LR2xbdn05** where only in image types (iii) and (iv) appears amongst the best three. Tables 2 to 7 also show that the other two best SR methods are **25FSR (Zhao et al. 2015)** and **07LSE (Freedman and R. 2011)**. Method **08SDS (Dong et al. 2011)** is a sparse coding type method, whereas **25FSR (Zhao et al. 2015)** is of PSF deconvolution type and **07LSE (Freedman and R. 2011)** can be considered as an hybrid method. As it was stated in Section 5.2, sparse coding methods are the focus of a considerable attention nowadays due to their promising results. On the other hand, PSF deconvolution methods also work relatively well but tend to be dependent on a good quality PSF kernel characterisation.

Other interesting results in Tables 2 to 7 should be highlighted: (1) none of the best three SR methods, according to the figures of merit with reference image, appear as the best three in terms of computational efficiency; (2) none of them appear as the best three when using figures of merit that do not need reference image (BRISQUE and NIQE); (3) method **12BDB (Michaeli and M. 2014)** is one of the best three SR methods when considering the BRISQUE and NIQE figures of merit. However, it is not among the first 10 best SR methods in terms of the figures of merit with reference. This fact is somehow surprising and confirms the need to continue working in the development of new and effective quality measures applied to super resolution.

24

In order to make these comparisons in a more meaningful and illustrative way, the data for each comparison case was normalised so that the interpolation method (**00IntL3**) for the lowest resolution and lowest noise value (in each comparison) acted as baseline measure (and therefore a value of zero is assigned to it), all methods that are better than **00IntL3** would get higher values than zero of the corresponding figure of merit, the best one would be set to 1 and the worst one a value equal to -1. The rest of the values of the remaining methods are linearly rescaled to $[-1, +1]$.

Figure 6(a)-(c) shows the bar plots for the SAM, RMSE and PSNR figures of merit for the comparison between a 2x resolution factor in relation to a 4x resolution factor. Notice that (as it may be expected) an increase in the resolution factor lowers down the reconstruction quality of all the methods. However, method **08SDS (Sparse Coding) (Dong et al. 2011)** gives the best result for the three figures of merit (SAM, RMSE and PSNR). Figure 6(b)-(c) also shows that the best three other methods for the RMSE and PSNR figures of merit are **25FSR (Zhao et al. 2015)**, **07LSE (Freedman and R. 2011)** and **03GPP (Gradient Profile) (Jian, Xu, and Shum 2008)**.

Figure 7(a)-(c) shows the effect of Gaussian noise addition on 2x super-resolution with different noise levels. In particular, the effect of the addition of $\sigma = 0.01$ and $\sigma = 0.05$ white Gaussian noise, respectively, is analysed. As we can see, the addition of noise lowers down the reconstruction results for all methods and all the three figures of merit (SAM, RMSE and PSNR). However, as it also happens in the results shown in Figure 6 the **08SDS (Sparse Coding) (Dong et al. 2011)** method provides the best results. Figure 7(b)-(c) also shows that the best three other methods for the RMSE and PSNR figures of merit the same as those of Figure 6, i.e. **25FSR (Zhao et al. 2015)**, **07LSE (Freedman and R. 2011)** and **03GPP (Gradient Profile) (Jian, Xu, and Shum 2008)**.

Figure 8(a)-(c) shows the effect of Gaussian noise addition on 4x super-resolution. In particular, the effect of the addition of $\sigma = 0.01$ and $\sigma = 0.05$ white Gaussian noise on the 4x super-resolution process is analysed. As we can see, the addition of noise decreases the super reconstruction results for all methods and the three figures of merit (SAM, RMSE and PSNR). However, again in the results shown in Figure 6 the **08SDS (Sparse Coding) (Dong et al. 2011)** method provides the best results. Figure 8(b) also shows that the best three other methods for the RMSE figures of merit are **25FSR (Zhao et al. 2015)**, **07LSE (Freedman and R. 2011)** and **04SRSI (Hybrid) (Glasner, S., and M. 2009)**, and methods **25FSR (Zhao et al. 2015)**, **07LSE (Freedman and R. 2011)** and **02FIU (PSF Deconvolution) (Shan et al. 2008)** for the PSNR (Figure 8(c)).

Figures 7 and 8 also show that the methods **15GLR (Timofte, V., and L. 2013)** to **19JOR (Dai, R., and L. 2015)** work relatively worse than the rest of methods (for the RMSE and PSNR figures of merit, above all).

Methods **10ANR (Timofte, V., and L. 2013)**, **15GLR (Timofte, V., and L. 2013)**, **16NELS (Timofte, V., and L. 2013)**, **17NENNLS (Timofte, V., and L. 2013)** and **18NELLE (H., T., and Y. 2004)** are variations of the same method where there is only a change in the regularisation factor, and therefore it may be logical that they have obtained similar performance results.

On the other hand, methods **11GPR (He and W.-C. 2011)**, **12BDB (Michaeli and M. 2014)**, **13SIP (Yang et al. 2008)**, **14SIS (Zeyde, M., and M. 2012)** and **19JOR) (Dai, R., and L. 2015)** are all different among them. However, results are similar and therefore model parameters might be playing a role in the results.

Codes related to the methods **10ANR (Timofte, V., and L. 2013)**, **15GLR (Timofte, V., and L. 2013)**, **16NELS (Timofte, V., and L. 2013)**, **17NENNLS (Timofte, V., and L. 2013)** and **18NELLE (H., T., and Y. 2004)** and **19JOR (Dai, R., and L. 2015)** were part of the same software toolbox and therefore the default values of the parameters used is shared among them. On the one hand, a 3×3 size patch with first and second order derivatives and Principal Component Analysis (PCA) as input for each vector was used in the toolbox for all of them. On the other hand, method **13SIP (Yang et al. 2008)** uses a 3×3 size patch and method **14SIS (Zeyde, M., and M. 2012)** and a dimensionality reduction process through PCA as well.

It seems clear that the ideal patch size depends on the selected methods and images. Each single method uses a different patch size. For instance, method **08SDS (Dong et al. 2011)** uses a 7×7 size patch, **06VSR (Yang et al. 2010)** and **25FSR (Zhao et al. 2015)** use a 5×5 size patch, **11GPR (He and W.-C. 2011)** uses a 20×20 size patch and method **22BSR (Polatkan et al. 2015)** an 8×8 size patch. We should bear in mind that we assume that there is no access to the HR image, there is no way to know the best value for the parameters of each one of the different methods. That is why the complete experimental part was made using the default parameters offered by authors of each method, which were assumed to work satisfactorily in the present case. The same patch size was used for those methods that used a dictionary training step, in order to make a fairer comparison.

Method **12BDB (Michaeli and M. 2014)** may have a low performance result because the kernel the method recovers is the one used in the interpolation stage and in this case the deblurring stage is made through an L_2 fitting minimising the number of patches.

Method **09DSF (Mapping) (Yang and M.-H. 2013)** is not shown in the plots of Figures 6 to 8 because performance results for this method were so poor it made visualisation of the rest of the methods performance results difficult to be obtained.

Figures 9 and 10 show some visual results on databases **01LR2xbd** and **02LR4xbd** in order to provide a qualitative super-resolution evaluation. Orthoimages are shown in rows and columns are arranged as follows: (1st) HR reference image, (2nd) baseline interpolation, (3rd) best PSNR reconstruction result, (4th) best PSNR learning result and (5th) best PSNR hybrid result.

When considering a 2x up-scaling factor (Fig. 9), method **08SDS (Dong et al. 2011)** is capable of recovering high-resolution details that are clearly missing in the rest of SR results. Structured details, like pedestrian crossings of images Fig. 9(i) and Fig. 9(n), are better-defined and visually closer to their HR counterparts. Other details where no regular structures can be detected, like shrubs appearing in images Fig. 9(d) and Fig. 9(s), are better-defined as well but sometimes it is possible to perceive a slight noise increase. This observation is also confirmed by quantitative evaluation where **08SDS (Dong et al. 2011)** PSNR values tend to be higher for (2nd) and (3rd) images than those for (1st) and (4th).

In the case of Fig. 10 (4x scaling factor), we can see how visual differences among SR results become relatively small. Even though **08SDS (Dong et al. 2011)** often seems to obtain the best result, other methods like **07LSE (Freedman and R. 2011)** and **26TSE (Huang, A., and N. 2015)** tend to generate slightly sharper edges and less noisy surfaces such as in the case of Fig. 10(e) and Fig. 10(o) images.
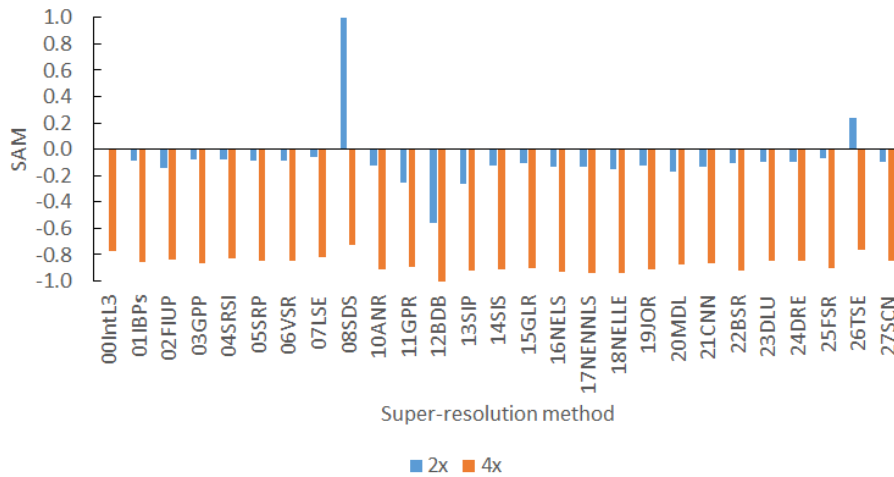
(a) Effect of super resolution factors using SAM metric.
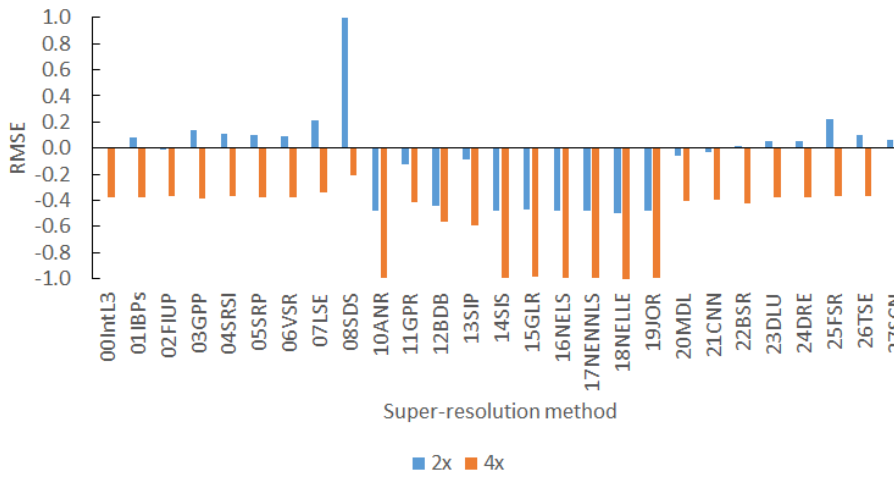
(b) Effect of super resolution factors using RMSE metric.
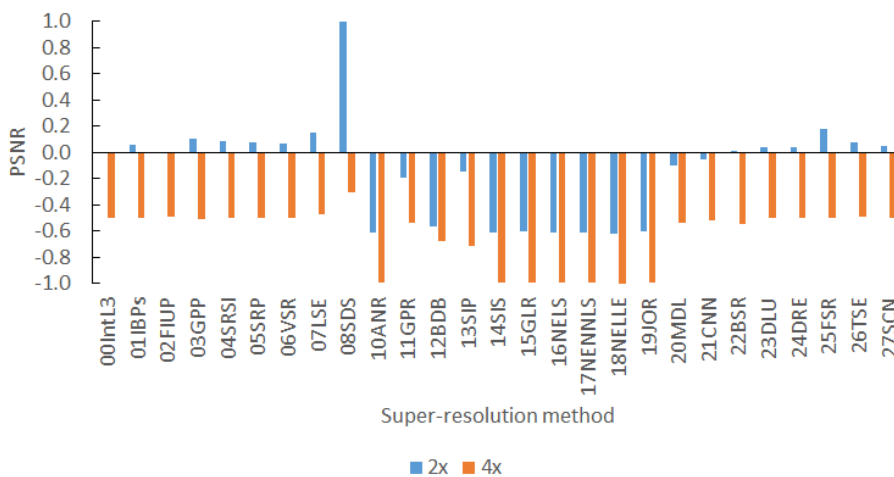
(c) Effect of super resolution factors using PSNR metric.

Figure 6.: Bar plots for 2x and 4x resolution factors without noise.
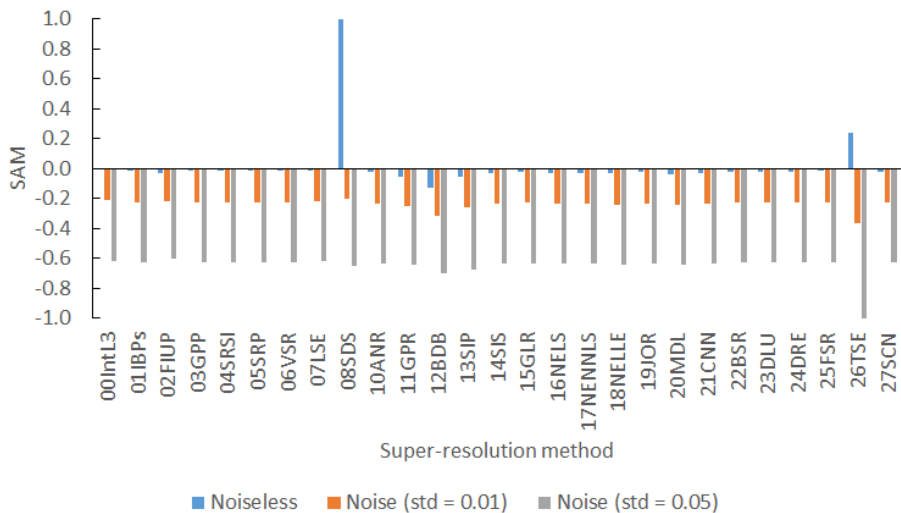
(a) Effect of noise on 2x super resolution factor using SAM metric.
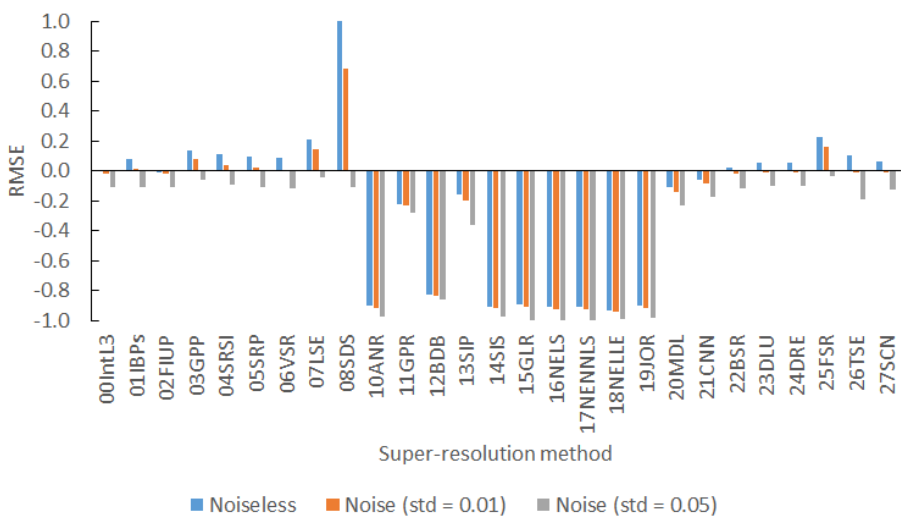
(b) Effect of noise on 2x super resolution factor using RMSE metric.
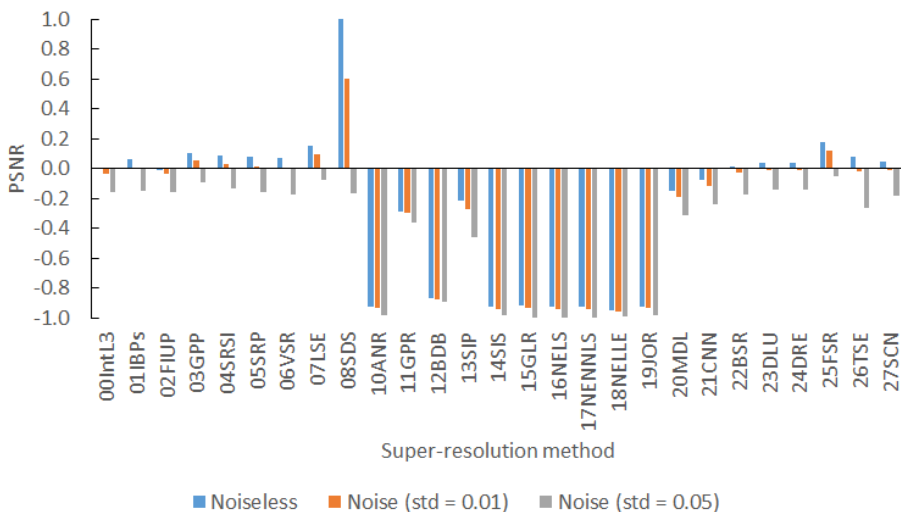
(c) Effect of noise on 2x super resolution factor using PSNR metric.

Figure 7.: Bar plots for 2x super resolution without and with $\sigma = 0.01$ and $\sigma = 0.05$ levels of additive white Gaussian noise.
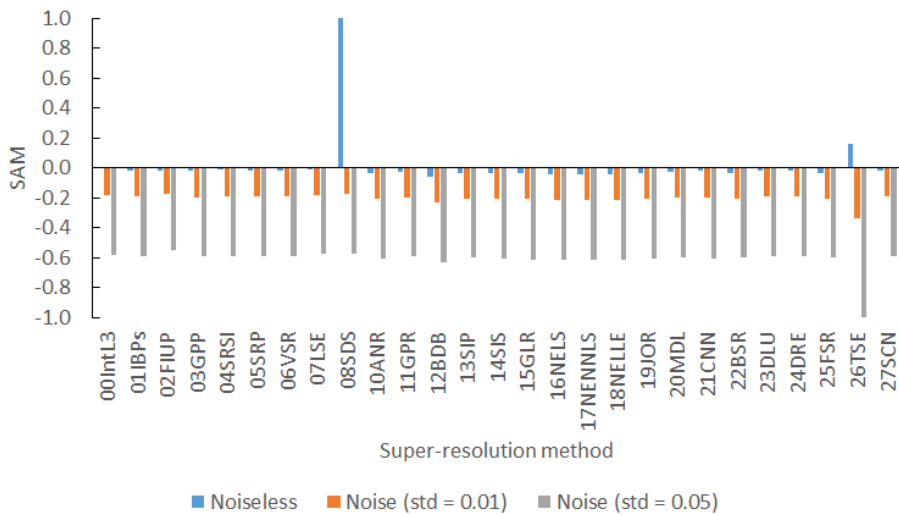
(a) Effect of noise on 4x super resolution factor using SAM metric.
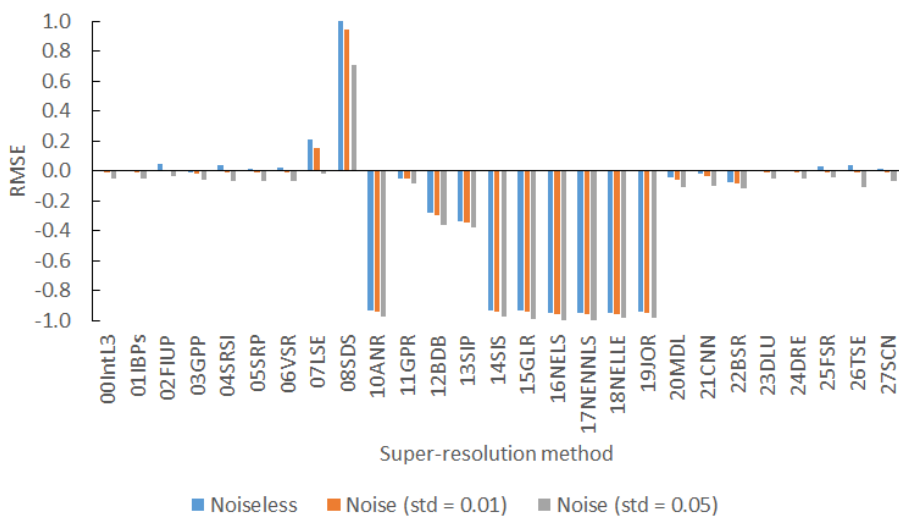
(b) Effect of noise on 4x super resolution factor using RMSE metric.
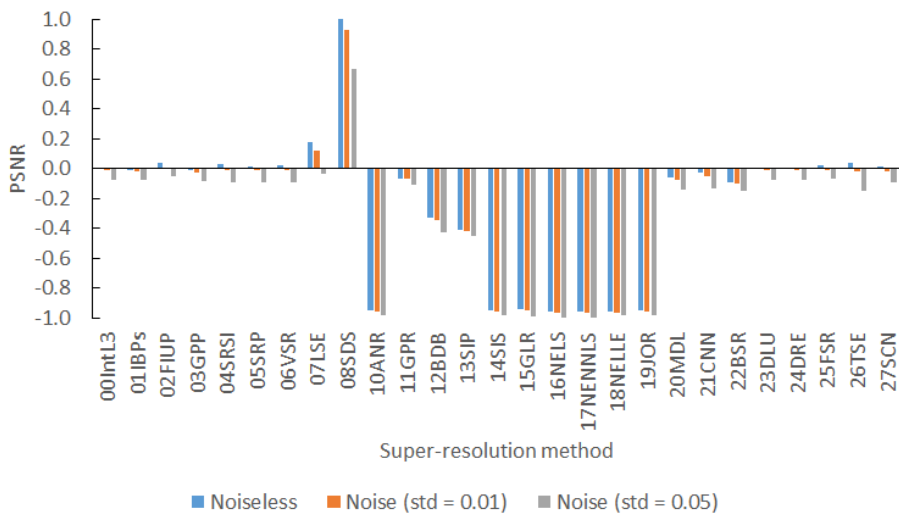
(c) Effect of noise on 4x super resolution factor using PSNR metric.

Figure 8.: Bar plots for 4x super resolution factor without and with $\sigma = 0.01$ and $\sigma = 0.05$ levels of additive white Gaussian noise.

Figure 9.: SR results for the (*a*) fold of **01LR2xbd** database. Orthoimages in rows and SR methods in columns. HR reference image in first column, Lanczos-3 interpolation in second, best PSNR reconstruction result in third, best PSNR learning result in forth and best PSNR hybrid result in fifth. For each method PSNR values in brackets and for each row the best PSNR result is highlighted in bold.

Table 2.: Average results of Time, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **01LR2x_bd** database. There are four group of rows (from (*i*) to (*iv*)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

01LR2x_bd	Id. & efficiency	Method identification code & metric with reference							Id. & metric without reference	
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(*i*)	**(00) 0.11 ± 0.03**	(00) 0.61 ± 0.16	(00) 9.70 ± 1.54	(00) 3.98 ± 0.74	(00) 0.9972 ± 0.0009	(00) 0.7904 ± 0.0419	(00) 0.8930 ± 0.0159	(00) 28.52 ± 1.42	(00) 41.26 ± 5.58	(00) 18.87 ± 1.16
	(24) 0.50 ± 0.14	(25) 0.64 ± 0.14	(25) 8.80 ± 1.19	(25) 3.60 ± 0.53	(25) 0.9978 ± 0.0006	(03) 0.8065 ± 0.0413	(01) 0.9079 ± 0.0163	(25) 29.34 ± 1.18	(01) 38.48 ± 5.26	(01) 15.25 ± 0.99
	(27) 1.50 ± 0.03	**(08) 0.54 ± 0.14**	**(08) 6.51 ± 1.50**	**(08) 2.66 ± 0.62**	**(08) 0.9987 ± 0.0005**	**(08) 0.8835 ± 0.0446**	**(08) 0.9580 ± 0.0196**	**(08) 32.09 ± 2.03**	(17) 29.63 ± 4.63	**(13) 8.56 ± 1.19**
	(12) 7.78 ± 8.02	(26) 0.55 ± 0.27	(07) 9.22 ± 1.19	(07) 3.73 ± 0.52	(07) 0.9976 ± 0.0006	(26) 0.8744 ± 0.0681	(07) 0.9220 ± 0.0140	(07) 28.93 ± 1.11	**(12) 21.17 ± 3.25**	(12) 11.55 ± 1.22
(*ii*)	**(00) 0.08 ± 0.02**	(00) 0.35 ± 0.18	(00) 9.49 ± 1.57	(00) 2.84 ± 0.76	(00) 0.9985 ± 0.0007	(00) 0.9009 ± 0.0250	(00) 0.9247 ± 0.0085	(00) 28.72 ± 1.44	(00) 39.48 ± 6.68	(00) 17.53 ± 2.48
	(24) 0.52 ± 0.09	(25) 0.38 ± 0.19	(25) 8.40 ± 1.28	(25) 2.51 ± 0.65	(25) 0.9988 ± 0.0006	(25) 0.9137 ± 0.0241	(25) 0.9448 ± 0.0071	(25) 29.75 ± 1.29	(01) 36.47 ± 7.01	(01) 14.88 ± 1.40
	(27) 1.52 ± 0.09	**(08) 0.34 ± 0.15**	**(08) 4.98 ± 0.45**	**(08) 1.48 ± 0.26**	**(08) 0.9996 ± 0.0001**	**(08) 0.9631 ± 0.0113**	**(08) 0.9853 ± 0.0019**	**(08) 34.25 ± 0.79**	(17) 26.94 ± 7.69	**(13) 8.13 ± 0.53**
	(12) 5.17 ± 3.46	(07) 0.38 ± 0.18	(04) 8.96 ± 1.44	(07) 2.65 ± 0.45	(07) 0.9987 ± 0.0004	(07) 0.9325 ± 0.0184	(07) 0.9563 ± 0.0051	(04) 29.20 ± 1.39	**(12) 20.75 ± 9.69**	(12) 11.14 ± 0.93
(*iii*)	**(00) 0.13 ± 0.03**	(00) 0.48 ± 0.04	(00) 11.97 ± 0.69	(00) 4.21 ± 0.33	(00) 0.9969 ± 0.0005	(00) 0.9323 ± 0.0126	(00) 0.9245 ± 0.0085	(00) 26.61 ± 0.52	(00) 32.19 ± 3.21	(00) 18.85 ± 0.64
	(24) 0.49 ± 0.14	(25) 0.53 ± 0.04	(25) 10.58 ± 0.54	(25) 3.72 ± 0.26	(25) 0.9976 ± 0.0003	(25) 0.9445 ± 0.0103	(25) 0.9467 ± 0.0056	(25) 27.69 ± 0.46	(01) 28.92 ± 3.43	(01) 16.08 ± 1.13
	(27) 1.51 ± 0.04	**(08) 0.48 ± 0.03**	**(08) 5.37 ± 0.25**	**(08) 1.89 ± 0.13**	**(08) 0.9994 ± 0.0001**	**(08) 0.9836 ± 0.0045**	**(08) 0.9892 ± 0.0018**	**(08) 33.57 ± 0.41**	(17) 18.35 ± 3.25	**(13) 7.69 ± 0.71**
	(12) 8.04 ± 7.28	(07) 0.53 ± 0.04	(07) 10.25 ± 0.41	(07) 3.56 ± 0.18	(07) 0.9978 ± 0.0002	(07) 0.9589 ± 0.0077	(07) 0.9579 ± 0.0043	(07) 27.95 ± 0.35	**(12) 7.73 ± 7.32**	(12) 12.14 ± 0.98
(*iv*)	**(00) 0.15 ± 0.12**	(00) 1.12 ± 0.36	(00) 11.75 ± 1.10	(00) 7.01 ± 1.40	(00) 0.9921 ± 0.0029	(00) 0.8991 ± 0.0174	(00) 0.9043 ± 0.0099	(00) 26.80 ± 0.82	(00) 43.50 ± 1.25	(00) 20.34 ± 0.82
	(24) 0.36 ± 0.04	(00) 1.15 ± 0.36	(25) 10.91 ± 0.99	(25) 6.52 ± 1.35	(25) 0.9932 ± 0.0026	(03) 0.9118 ± 0.0156	(25) 0.9263 ± 0.0090	(25) 27.44 ± 0.80	(01) 39.78 ± 1.52	(01) 16.38 ± 0.95
	(27) 1.49 ± 0.03	(08) 0.94 ± 0.29	**(08) 7.19 ± 0.66**	**(08) 4.31 ± 1.01**	**(08) 0.9970 ± 0.0013**	**(08) 0.9648 ± 0.0082**	**(08) 0.9823 ± 0.0032**	**(08) 31.06 ± 0.78**	(08) 30.36 ± 0.91	**(13) 10.19 ± 0.50**
	(12) 4.75 ± 0.74	**(26) 0.70 ± 0.46**	(26) 10.17 ± 2.42	(26) 4.80 ± 2.63	(26) 0.9954 ± 0.0043	(07) 0.9296 ± 0.0135	(07) 0.9423 ± 0.0069	(26) 28.23 ± 2.18	(26) 31.98 ± 4.99	(12) 11.95 ± 0.39

30

(a) HR (b) 00IntL3 (25.83) (c) 02FIU (26.02) (d) 08SDS (27.14) (e) 07LSE (26.04)

(f) HR (g) 00IntL3 (22.07) (h) 25FSR (22.43) (i) 08SDS (23.99) (j) 07LSE (22.41)

(k) HR (l) 00IntL3 (20.89) (m) 02FIU (21.10) (n) 08SDS (22.83) (o) 26TSE (23.82)

(p) HR (q) 00IntL3 (21.71) (r) 02FIU (21.75) (s) 08SDS (23.26) (t) 26TSE (21.76)
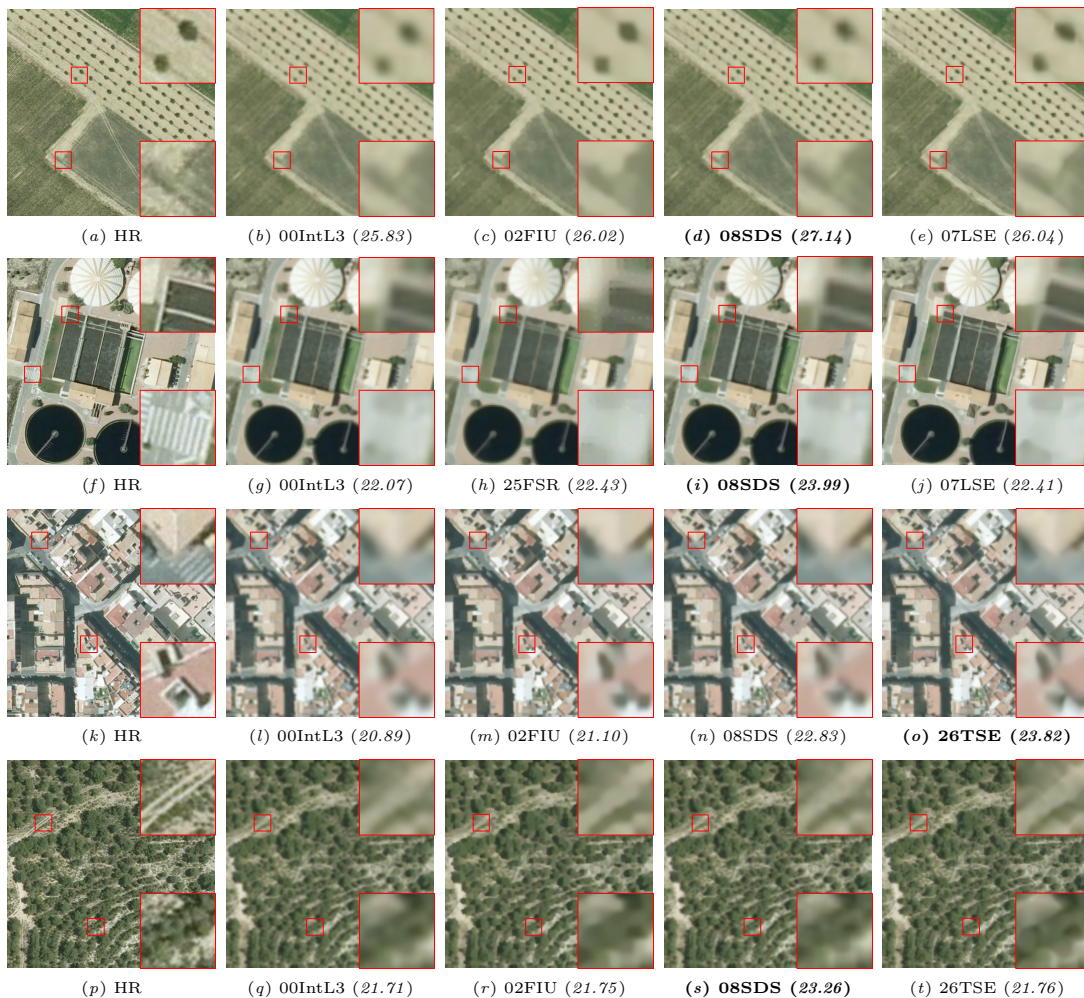
Figure 10.: SR results for the (e) fold of **02LR4xbd** database. Orthoimages in rows and SR methods in columns. HR reference image in first column, Lanczos-3 interpolation in second, best PSNR reconstruction result in third, best PSNR learning result in forth and best PSNR hybrid result in fifth. For each method PSNR values in brackets and for each row the best PSNR result is highlighted in bold.

Table 3.: Average results of TIME, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **02LR4x_bd** database. There are four group of rows (from (i) to (iv)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

02LR4x_bd	Id. & efficiency	Method identification code & metric with reference							Id. & metric without reference	
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(i)	(00) 0.07 ± 0.02	(00) 0.95 ± 0.27	(00) 15.83 ± 2.81	(00) 3.27 ± 0.75	(00) 0.9925 ± 0.0032	(00) 0.5687 ± 0.0585	(00) 0.6779 ± 0.0176	(00) 24.29 ± 1.60	(00) 66.42 ± 5.94	(00) 39.11 ± 4.55
	(24) 0.61 ± 0.12	(23) 0.99 ± 0.27	(25) 15.79 ± 2.55	(25) 3.24 ± 0.71	(25) 0.9926 ± 0.0030	(23) 0.5703 ± 0.0585	(23) 0.6788 ± 0.0175	(02) 24.30 ± 1.67	(02) 63.31 ± 6.25	(02) 32.71 ± 4.39
	(15) 1.16 ± 0.22	(08) 0.91 ± 0.23	(08) 12.46 ± 1.47	(08) 2.55 ± 0.39	(08) 0.9955 ± 0.0012	(08) 0.6527 ± 0.0430	(08) 0.7607 ± 0.0183	(08) 26.31 ± 1.05	(13) 34.61 ± 8.66	(17) 15.27 ± 2.32
	(12) 13.37 ± 11.39	(26) 0.94 ± 0.45	(07) 15.10 ± 2.19	(07) 3.05 ± 0.58	(07) 0.9934 ± 0.0023	(26) 0.6707 ± 0.0746	(07) 0.7219 ± 0.0144	(07) 24.66 ± 1.30	(12) 29.91 ± 3.25	(12) 16.70 ± 1.20
(ii)	(00) 0.08 ± 0.03	(00) 0.62 ± 0.29	(00) 16.82 ± 2.98	(00) 2.52 ± 0.67	(00) 0.9951 ± 0.0023	(00) 0.7120 ± 0.0623	(00) 0.7028 ± 0.0251	(00) 23.76 ± 1.59	(00) 65.60 ± 2.94	(00) 36.08 ± 5.18
	(24) 0.60 ± 0.19	(25) 0.65 ± 0.30	(25) 16.63 ± 3.16	(25) 2.48 ± 0.66	(25) 0.9953 ± 0.0022	(02) 0.7167 ± 0.0646	(02) 0.7076 ± 0.0257	(25) 23.88 ± 1.72	(02) 60.35 ± 4.51	(02) 30.81 ± 4.78
	(15) 1.26 ± 0.33	(08) 0.61 ± 0.29	(08) 13.62 ± 2.36	(08) 2.04 ± 0.53	(08) 0.9968 ± 0.0014	(08) 0.7918 ± 0.0524	(08) 0.7905 ± 0.0213	(08) 25.59 ± 1.55	(13) 28.57 ± 6.53	(17) 13.19 ± 2.13
	(12) 12.01 ± 10.75	(07) 0.65 ± 0.29	(07) 16.52 ± 2.46	(07) 2.43 ± 0.58	(07) 0.9955 ± 0.0019	(07) 0.7536 ± 0.0576	(07) 0.7486 ± 0.0235	(07) 23.89 ± 1.32	(12) 32.53 ± 8.09	(12) 14.83 ± 1.86
(iii)	(00) 0.08 ± 0.03	(00) 0.90 ± 0.09	(00) 21.82 ± 1.24	(00) 3.83 ± 0.26	(00) 0.9897 ± 0.0013	(00) 0.7682 ± 0.0347	(00) 0.6861 ± 0.0258	(00) 21.40 ± 0.51	(00) 60.71 ± 2.60	(00) 35.06 ± 1.84
	(24) 0.47 ± 0.14	(02) 0.93 ± 0.09	(25) 21.43 ± 1.31	(02) 3.77 ± 0.28	(02) 0.9901 ± 0.0014	(02) 0.7816 ± 0.0359	(02) 0.7015 ± 0.0284	(02) 21.56 ± 0.55	(02) 52.15 ± 3.84	(02) 33.03 ± 1.07
	(15) 1.25 ± 0.22	(08) 0.91 ± 0.09	(08) 17.43 ± 0.92	(08) 3.06 ± 0.18	(08) 0.9935 ± 0.0008	(08) 0.8440 ± 0.0240	(08) 0.7859 ± 0.0158	(08) 23.35 ± 0.47	(13) 19.08 ± 1.55	(17) 14.41 ± 2.69
	(12) 13.70 ± 12.05	(04) 0.93 ± 0.09	(26) 19.64 ± 2.65	(07) 3.58 ± 0.24	(07) 0.9909 ± 0.0012	(07) 0.8089 ± 0.0307	(07) 0.7387 ± 0.0222	(26) 22.36 ± 1.17	(12) 23.80 ± 0.99	(12) 17.24 ± 2.62
(iv)	(00) 0.08 ± 0.02	(00) 1.76 ± 0.50	(00) 19.13 ± 1.93	(00) 5.67 ± 0.94	(00) 0.9793 ± 0.0060	(00) 0.6982 ± 0.0339	(00) 0.6208 ± 0.0225	(00) 22.57 ± 0.89	(00) 68.11 ± 1.02	(00) 46.66 ± 0.43
	(24) 0.50 ± 0.16	(02) 1.79 ± 0.50	(02) 19.06 ± 1.92	(02) 5.66 ± 0.94	(02) 0.9794 ± 0.0060	(02) 0.7083 ± 0.0339	(02) 0.6287 ± 0.0227	(02) 22.60 ± 0.89	(02) 53.32 ± 2.50	(02) 42.03 ± 2.08
	(15) 1.22 ± 0.15	(08) 1.71 ± 0.49	(08) 16.07 ± 1.49	(08) 4.79 ± 0.88	(08) 0.9853 ± 0.0048	(08) 0.7822 ± 0.0308	(08) 0.7284 ± 0.0207	(08) 24.07 ± 0.82	(13) 33.31 ± 2.87	(17) 21.42 ± 0.93
	(12) 12.87 ± 10.44	(26) 1.15 ± 0.68	(26) 17.49 ± 4.42	(26) 4.07 ± 2.07	(26) 0.9870 ± 0.0110	(07) 0.7459 ± 0.0301	(07) 0.6784 ± 0.0211	(26) 23.55 ± 2.33	(12) 34.11 ± 1.95	(12) 20.61 ± 1.34

Table 4.: Average results of Time, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **03LR2x_bdn01** database. There are four group of rows (from (i) to (iv)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

03LR2x_bdn01	Id. & efficiency	Method identification code & metric with reference							Id. & metric without reference	
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(i)	(00) 0.12 ± 0.07	(00) 1.10 ± 0.17	(00) 9.97 ± 1.49	(00) 4.09 ± 0.73	(00) 0.9971 ± 0.0009	(00) 0.7739 ± 0.0424	(00) 0.8733 ± 0.0137	(00) 28.27 ± 1.33	(00) 39.92 ± 3.19	(00) 19.01 ± 1.08
	(24) 0.47 ± 0.07	(02) 1.12 ± 0.18	(25) 9.14 ± 1.15	(25) 3.74 ± 0.53	(25) 0.9976 ± 0.0006	(03) 0.7880 ± 0.0419	(25) 0.8868 ± 0.0120	(25) 29.00 ± 1.09	(23) 37.91 ± 2.50	(01) 14.45 ± 0.54
	(27) 1.48 ± 0.02	(08) 1.09 ± 0.16	(08) 7.91 ± 1.24	(08) 3.22 ± 0.55	(08) 0.9982 ± 0.0006	(08) 0.8324 ± 0.0458	(08) 0.9118 ± 0.0173	(08) 30.29 ± 1.37	(17) 29.11 ± 2.56	(13) 7.94 ± 0.82
	(12) 7.97 ± 6.59	(07) 1.13 ± 0.16	(07) 9.57 ± 1.14	(07) 3.87 ± 0.51	(07) 0.9974 ± 0.0006	(26) 0.8553 ± 0.0748	(07) 0.8997 ± 0.0119	(07) 28.59 ± 1.02	(12) 20.23 ± 4.32	(12) 11.65 ± 0.55
(ii)	(00) 0.07 ± 0.02	(00) 0.76 ± 0.26	(00) 9.76 ± 1.53	(00) 2.92 ± 0.75	(00) 0.9984 ± 0.0007	(00) 0.8907 ± 0.0280	(00) 0.9065 ± 0.0101	(00) 28.46 ± 1.36	(00) 38.66 ± 6.20	(00) 18.35 ± 1.67
	(24) 0.48 ± 0.05	(02) 0.76 ± 0.27	(25) 8.77 ± 1.23	(25) 2.62 ± 0.65	(25) 0.9987 ± 0.0006	(25) 0.9016 ± 0.0277	(25) 0.9263 ± 0.0089	(25) 29.37 ± 1.20	(24) 36.45 ± 5.72	(01) 14.51 ± 0.83
	(27) 1.47 ± 0.02	(08) 0.76 ± 0.25	(08) 6.65 ± 0.34	(08) 1.97 ± 0.29	(08) 0.9993 ± 0.0002	(08) 0.9303 ± 0.0202	(08) 0.9421 ± 0.0060	(08) 31.72 ± 0.45	(17) 25.82 ± 5.49	(13) 6.81 ± 0.73
	(12) 13.42 ± 9.22	(07) 0.78 ± 0.26	(07) 9.34 ± 0.67	(07) 2.75 ± 0.45	(07) 0.9986 ± 0.0004	(07) 0.9202 ± 0.0220	(07) 0.9365 ± 0.0069	(04) 28.83 ± 1.28	(12) 20.20 ± 7.64	(12) 10.91 ± 1.16
(iii)	(00) 0.10 ± 0.07	(00) 0.96 ± 0.05	(00) 12.19 ± 0.68	(00) 4.28 ± 0.33	(00) 0.9968 ± 0.0005	(00) 0.9282 ± 0.0137	(00) 0.9117 ± 0.0089	(00) 26.46 ± 0.50	(00) 33.58 ± 2.44	(00) 18.86 ± 0.75
	(24) 0.56 ± 0.14	(02) 0.96 ± 0.05	(25) 10.87 ± 0.53	(25) 3.82 ± 0.25	(25) 0.9975 ± 0.0003	(25) 0.9396 ± 0.0116	(25) 0.9337 ± 0.0064	(25) 27.45 ± 0.43	(01) 31.05 ± 2.99	(01) 15.44 ± 0.89
	(27) 1.47 ± 0.01	(08) 0.97 ± 0.05	(08) 7.02 ± 0.20	(08) 2.47 ± 0.14	(08) 0.9989 ± 0.0001	(08) 0.9696 ± 0.0085	(08) 0.9587 ± 0.0061	(08) 31.24 ± 0.25	(17) 19.27 ± 2.63	(13) 6.18 ± 0.48
	(12) 10.15 ± 9.09	(07) 0.98 ± 0.05	(07) 10.55 ± 0.39	(07) 3.67 ± 0.19	(07) 0.9976 ± 0.0002	(07) 0.9540 ± 0.0093	(07) 0.9441 ± 0.0057	(07) 27.70 ± 0.33	(12) 10.91 ± 4.96	(12) 12.63 ± 1.66
(iv)	(00) 0.07 ± 0.01	(00) 1.76 ± 0.46	(00) 11.97 ± 1.08	(00) 7.14 ± 1.42	(00) 0.9918 ± 0.0029	(00) 0.8952 ± 0.0179	(00) 0.8969 ± 0.0105	(00) 26.63 ± 0.79	(00) 42.78 ± 1.08	(00) 20.07 ± 0.87
	(24) 0.47 ± 0.12	(02) 1.78 ± 0.48	(25) 11.18 ± 0.97	(25) 6.68 ± 1.38	(25) 0.9928 ± 0.0027	(03) 0.9074 ± 0.0161	(25) 0.9185 ± 0.0098	(25) 27.22 ± 0.76	(01) 39.37 ± 1.36	(01) 15.94 ± 0.79
	(27) 1.47 ± 0.01	(08) 1.70 ± 0.44	(08) 8.48 ± 0.57	(08) 5.07 ± 1.10	(08) 0.9958 ± 0.0016	(08) 0.9505 ± 0.0104	(08) 0.9642 ± 0.0054	(08) 29.61 ± 0.58	(08) 28.68 ± 0.40	(13) 9.17 ± 0.47
	(12) 14.25 ± 9.68	(26) 1.64 ± 0.76	(26) 10.76 ± 2.27	(26) 5.06 ± 2.67	(26) 0.9950 ± 0.0046	(07) 0.9246 ± 0.0142	(07) 0.9340 ± 0.0077	(07) 27.70 ± 1.92	(12) 32.11 ± 3.50	(12) 11.91 ± 0.86

Table 5.: Average results of Time, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **04LR4x_bdn01** database. There are four group of rows (from (i) to (iv)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

04LR4x_bdn01	Id. & efficiency	Method identification code & metric with reference							Id. & metric without reference	
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(i)	(00) 0.09 ± 0.03	(00) 1.36 ± 0.28	(00) 15.99 ± 2.77	(00) 3.30 ± 0.75	(00) 0.9924 ± 0.0032	(00) 0.5578 ± 0.0596	(00) 0.6618 ± 0.0153	(00) 24.20 ± 1.56	(00) 64.57 ± 4.50	(00) 39.43 ± 2.83
	(24) 0.39 ± 0.06	(02) 1.36 ± 0.29	(02) 15.96 ± 2.89	(25) 3.28 ± 0.71	(25) 0.9924 ± 0.0031	(23) 0.5587 ± 0.0597	(02) 0.6628 ± 0.0133	(02) 24.22 ± 1.63	(03) 61.60 ± 3.75	(02) 33.50 ± 3.22
	(15) 1.07 ± 0.30	(08) 1.32 ± 0.24	(08) 12.71 ± 1.43	(08) 2.61 ± 0.38	(08) 0.9953 ± 0.0012	(08) 0.6386 ± 0.0450	(08) 0.7411 ± 0.0157	(08) 26.13 ± 1.00	(13) 34.35 ± 8.46	(17) 18.87 ± 1.56
	(12) 9.02 ± 7.93	(07) 1.37 ± 0.27	(07) 15.33 ± 2.15	(07) 3.10 ± 0.58	(07) 0.9932 ± 0.0023	(26) 0.6560 ± 0.0799	(07) 0.7012 ± 0.0121	(07) 24.52 ± 1.26	(12) 42.50 ± 28.63	(12) 22.63 ± 3.81
(ii)	(00) 0.08 ± 0.02	(00) 0.95 ± 0.35	(00) 16.98 ± 2.95	(00) 2.54 ± 0.67	(00) 0.9951 ± 0.0023	(00) 0.7039 ± 0.0640	(00) 0.6871 ± 0.0260	(00) 23.68 ± 1.56	(00) 64.28 ± 2.80	(00) 35.77 ± 3.69
	(24) 0.39 ± 0.06	(02) 0.94 ± 0.35	(02) 16.81 ± 2.86	(25) 2.51 ± 0.65	(25) 0.9952 ± 0.0022	(02) 0.7108 ± 0.0654	(02) 0.6959 ± 0.0262	(25) 23.78 ± 1.68	(02) 58.45 ± 4.75	(02) 32.23 ± 5.31
	(15) 0.94 ± 0.19	(08) 0.94 ± 0.34	(08) 13.84 ± 2.32	(08) 2.07 ± 0.53	(08) 0.9967 ± 0.0014	(08) 0.7815 ± 0.0545	(08) 0.7713 ± 0.0220	(08) 25.44 ± 1.50	(13) 29.44 ± 7.15	(17) 15.62 ± 1.86
	(12) 8.97 ± 7.76	(07) 0.96 ± 0.35	(07) 16.72 ± 2.44	(07) 2.46 ± 0.58	(07) 0.9954 ± 0.0020	(07) 0.7438 ± 0.0594	(07) 0.7304 ± 0.0249	(07) 23.78 ± 1.29	(12) 31.51 ± 6.66	(12) 18.57 ± 1.34
(iii)	(00) 0.09 ± 0.05	(00) 1.25 ± 0.08	(00) 21.94 ± 1.24	(00) 3.86 ± 0.26	(00) 0.9896 ± 0.0013	(00) 0.7648 ± 0.0353	(00) 0.6755 ± 0.0252	(00) 21.35 ± 0.50	(00) 60.44 ± 2.23	(00) 36.78 ± 1.49
	(24) 0.41 ± 0.10	(02) 1.25 ± 0.09	(02) 21.55 ± 1.30	(02) 3.79 ± 0.28	(02) 0.9900 ± 0.0014	(02) 0.7787 ± 0.0360	(02) 0.6929 ± 0.0278	(02) 21.51 ± 0.54	(02) 51.44 ± 3.69	(02) 33.57 ± 0.89
	(15) 0.94 ± 0.14	(08) 1.26 ± 0.09	(08) 17.60 ± 0.92	(08) 3.09 ± 0.18	(08) 0.9934 ± 0.0008	(08) 0.8397 ± 0.0247	(08) 0.7725 ± 0.0153	(08) 23.27 ± 0.46	(13) 20.58 ± 1.63	(17) 14.98 ± 1.21
	(12) 1.27 ± 0.09	(07) 1.26 ± 0.09	(26) 19.97 ± 2.61	(07) 3.60 ± 0.24	(07) 0.9907 ± 0.0012	(07) 0.8043 ± 0.0319	(07) 0.7264 ± 0.0220	(26) 22.22 ± 1.14	(12) 26.77 ± 1.31	(12) 20.29 ± 0.69
(iv)	(00) 0.08 ± 0.05	(00) 2.22 ± 0.57	(00) 19.27 ± 1.91	(00) 5.71 ± 0.95	(00) 0.9789 ± 0.0061	(00) 0.6947 ± 0.0345	(00) 0.6147 ± 0.0227	(00) 22.25 ± 0.88	(00) 66.81 ± 0.92	(00) 45.02 ± 1.73
	(24) 0.45 ± 0.11	(02) 2.20 ± 0.57	(02) 19.19 ± 1.90	(02) 5.69 ± 0.95	(02) 0.9791 ± 0.0061	(02) 0.7053 ± 0.0345	(02) 0.6237 ± 0.0230	(02) 22.54 ± 0.88	(02) 53.45 ± 2.70	(02) 41.73 ± 2.71
	(15) 0.90 ± 0.11	(08) 2.18 ± 0.57	(08) 16.27 ± 1.47	(08) 4.85 ± 0.89	(08) 0.9849 ± 0.0049	(08) 0.7776 ± 0.0314	(08) 0.7203 ± 0.0209	(08) 23.96 ± 0.80	(13) 32.91 ± 2.95	(17) 19.87 ± 1.44
	(12) 9.91 ± 8.39	(26) 1.92 ± 0.90	(26) 17.87 ± 4.32	(26) 4.15 ± 2.08	(26) 0.9866 ± 0.0113	(07) 0.7409 ± 0.0308	(07) 0.6704 ± 0.0214	(26) 23.34 ± 2.23	(12) 33.69 ± 1.88	(12) 20.90 ± 1.62

Table 6.: Average results of Time, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **05LR2x_bdn05** database. There are four group of rows (from (i) to (iv)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

05LR2x_bdn05	Id. & efficiency	Method identification code & metric with reference							Id. & metric without reference	
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(i)	(00) 0.08 ± 0.05	(00) 2.05 ± 0.24	(00) 10.99 ± 1.35	(00) 4.50 ± 0.70	(00) 0.9965 ± 0.0010	(00) 0.7266 ± 0.0459	(00) 0.8084 ± 0.0107	(00) 27.40 ± 1.09	(00) 37.32 ± 1.28	(00) 16.65 ± 0.83
	(24) 0.37 ± 0.08	(02) 2.00 ± 0.25	(25) 10.31 ± 1.03	(25) 4.22 ± 0.51	(25) 0.9969 ± 0.0007	(03) 0.7383 ± 0.0454	(25) 0.8232 ± 0.0102	(25) 27.94 ± 0.87	(24) 38.43 ± 2.30	(01) 11.75 ± 0.46
	(27) 1.46 ± 0.01	(06) 2.06 ± 0.25	(22) 11.18 ± 1.36	(22) 4.57 ± 0.71	(22) 0.9964 ± 0.0010	(05) 0.7232 ± 0.0466	(18) 0.8043 ± 0.0112	(22) 27.25 ± 1.08	(20) 34.13 ± 3.23	(13) 7.17 ± 0.38
	(12) 8.38 ± 6.26	(07) 2.06 ± 0.24	(07) 10.82 ± 1.01	(07) 4.37 ± 0.49	(07) 0.9967 ± 0.0007	(26) 0.8012 ± 0.0915	(07) 0.8280 ± 0.0115	(07) 27.51 ± 0.80	(12) 19.72 ± 4.55	(12) 10.34 ± 0.38
(ii)	(00) 0.08 ± 0.03	(00) 1.48 ± 0.43	(00) 10.79 ± 1.38	(00) 3.22 ± 0.74	(00) 0.9980 ± 0.0008	(00) 0.8581 ± 0.0359	(00) 0.8458 ± 0.0147	(00) 27.56 ± 1.10	(00) 35.88 ± 3.88	(00) 16.68 ± 0.70
	(24) 0.38 ± 0.04	(02) 1.45 ± 0.43	(25) 9.99 ± 1.09	(25) 2.98 ± 0.64	(25) 0.9983 ± 0.0006	(02) 0.8717 ± 0.0325	(25) 0.8663 ± 0.0140	(25) 28.22 ± 0.93	(23) 35.81 ± 3.29	(01) 11.41 ± 0.62
	(27) 1.51 ± 0.06	(06) 1.49 ± 0.44	(05) 10.88 ± 1.17	(05) 3.24 ± 0.67	(05) 0.9980 ± 0.0007	(05) 0.8536 ± 0.0370	(18) 0.8396 ± 0.0153	(05) 27.47 ± 0.93	(17) 31.23 ± 1.14	(13) 5.94 ± 0.98
	(12) 9.07 ± 6.67	(07) 1.48 ± 0.42	(07) 10.57 ± 0.57	(07) 3.11 ± 0.47	(07) 0.9982 ± 0.0004	(07) 0.8821 ± 0.0320	(07) 0.8706 ± 0.0127	(07) 27.69 ± 0.47	(12) 19.19 ± 4.70	(12) 8.97 ± 0.79
(iii)	(00) 0.07 ± 0.03	(00) 1.84 ± 0.08	(00) 13.01 ± 0.63	(00) 4.57 ± 0.32	(00) 0.9964 ± 0.0005	(00) 0.9141 ± 0.0168	(00) 0.8692 ± 0.0120	(00) 25.89 ± 0.43	(00) 31.04 ± 2.11	(00) 17.36 ± 0.54
	(24) 0.33 ± 0.09	(02) 1.80 ± 0.08	(25) 11.88 ± 0.48	(25) 4.17 ± 0.25	(25) 0.9970 ± 0.0004	(02) 0.9249 ± 0.0150	(25) 0.8913 ± 0.0107	(25) 26.68 ± 0.36	(24) 29.59 ± 2.55	(01) 11.93 ± 0.27
	(27) 1.52 ± 0.05	(06) 1.86 ± 0.09	(08) 11.66 ± 0.15	(08) 4.09 ± 0.19	(08) 0.9971 ± 0.0003	(08) 0.9231 ± 0.0174	(14) 0.8662 ± 0.0125	(08) 26.83 ± 0.11	(17) 23.00 ± 3.38	(13) 5.51 ± 0.36
	(12) 11.29 ± 8.32	(07) 1.84 ± 0.08	(07) 11.65 ± 0.36	(07) 4.05 ± 0.19	(07) 0.9971 ± 0.0003	(07) 0.9374 ± 0.0134	(07) 0.8981 ± 0.0114	(07) 26.84 ± 0.27	(12) 12.71 ± 2.13	(12) 9.95 ± 0.74
(iv)	(00) 0.06 ± 0.01	(00) 3.12 ± 0.71	(00) 12.81 ± 1.01	(00) 7.65 ± 1.49	(00) 0.9906 ± 0.0033	(00) 0.8799 ± 0.0202	(00) 0.8695 ± 0.0132	(00) 26.03 ± 0.69	(00) 40.62 ± 1.02	(00) 18.08 ± 1.12
	(24) 0.30 ± 0.03	(02) 3.09 ± 0.72	(25) 12.14 ± 0.89	(25) 7.27 ± 1.46	(25) 0.9915 ± 0.0030	(02) 0.8920 ± 0.0176	(25) 0.8902 ± 0.0129	(25) 26.50 ± 0.64	(24) 38.01 ± 1.03	(01) 13.34 ± 0.49
	(27) 1.48 ± 0.01	(19) 3.13 ± 0.72	(08) 12.53 ± 0.38	(08) 7.45 ± 1.47	(08) 0.9910 ± 0.0031	(08) 0.8957 ± 0.0193	(08) 0.8982 ± 0.0135	(08) 26.21 ± 0.26	(17) 27.76 ± 1.83	(13) 7.44 ± 0.43
	(12) 8.53 ± 7.38	(07) 3.12 ± 0.71	(07) 11.91 ± 0.76	(26) 5.93 ± 2.85	(26) 0.9933 ± 0.0057	(07) 0.9062 ± 0.0173	(07) 0.9037 ± 0.0114	(07) 26.66 ± 0.56	(12) 32.34 ± 2.63	(12) 10.88 ± 0.63

Table 7.: Average results of Time, SAM, RMSE, sCC, Q-index, PSNR, BRISQUE and NIQE on **06LR4x_bdn05** database. There are four group of rows (from (*i*) to (*iv*)) which correspond to the four different kinds of orthoimages. Besides, each group contains four rows with the following results: baseline Lanczos-3 interpolation (first row), best reconstruction result (second row), best learning result (third row) and best hybrid result (fourth row). Note that the best results are highlighted in bold and method's identification codes are in brackets within the table.

| 06LR4x_bdn05 | Id. & efficiency | Method identification code & metric with reference | | | | | | | Id. & metric without reference | |
	Time (s)	SAM	RMSE	ERGAS	sCC	Q-index	SSIM	PSNR (dB)	BRISQUE	NIQE
(i)	(00) 0.07 ± 0.02	(00) 2.22 ± 0.32	(00) 16.66 ± 2.64	(00) 3.43 ± 0.73	(00) 0.9918 ± 0.0032	(00) 0.5264 ± 0.0623	(00) 0.6073 ± 0.0110	(00) 23.82 ± 1.42	(00) 61.73 ± 2.73	(00) 33.85 ± 1.80
	(24) 0.36 ± 0.08	**(02) 2.16 ± 0.32**	(02) 16.57 ± 2.77	(25) 3.42 ± 0.69	(25) 0.9919 ± 0.0031	(02) 0.5319 ± 0.0639	(02) 0.6209 ± 0.0095	(02) 23.88 ± 1.50	(03) 57.89 ± 1.58	(01) 31.01 ± 1.24
	(15) 0.95 ± 0.21	(08) 2.20 ± 0.30	**(08) 13.66 ± 1.31**	**(08) 2.80 ± 0.37**	**(08) 0.9946 ± 0.0013**	(08) 0.5978 ± 0.0497	**(08) 0.6731 ± 0.0163**	**(08) 25.49 ± 0.85**	**(13) 35.97 ± 6.50**	**(17) 16.15 ± 1.02**
	(12) 14.13 ± 2.44	(07) 2.20 ± 0.31	(07) 16.21 ± 2.01	(07) 3.27 ± 0.56	(07) 0.9925 ± 0.0024	**(26) 0.6142 ± 0.0917**	(07) 0.6330 ± 0.0127	(07) 24.02 ± 1.10	(12) 42.68 ± 24.07	(12) 19.67 ± 1.47
(ii)	(00) 0.06 ± 0.01	(00) 1.60 ± 0.60	(00) 17.60 ± 6.10	(00) 2.63 ± 7.58	(00) 0.9947 ± 0.9961	(00) 0.6777 ± 0.0975	(00) 0.6358 ± 0.0432	(00) 23.35 ± 10.68	(00) 61.25 ± 16.13	(00) 34.00 ± 11.44
	(24) 0.31 ± 0.04	**(02) 1.55 ± 0.60**	(02) 17.39 ± 6.07	(25) 2.60 ± 7.60	(02) 0.9948 ± 0.9807	(02) 0.6896 ± 0.0938	(02) 0.6538 ± 0.0451	(25) 23.46 ± 10.89	(02) 56.06 ± 12.38	(01) 32.73 ± 12.10
	(15) 1.00 ± 0.18	(08) 1.59 ± 460.81	**(08) 14.70 ± 4.91**	**(08) 2.20 ± 6.25**	**(08) 0.9963 ± 0.7471**	**(08) 0.7484 ± 0.0727**	**(08) 0.7067 ± 0.0412**	**(08) 24.89 ± 11.39**	**(13) 31.53 ± 1.39**	**(17) 14.85 ± 10.08**
	(12) 15.89 ± 4.03	(07) 1.59 ± 1331.10	(07) 17.50 ± 6.16	(07) 2.57 ± 7.36	(07) 0.9950 ± 0.9167	(07) 0.7110 ± 0.0865	(07) 0.6677 ± 0.0437	(07) 23.37 ± 10.63	(12) 32.92 ± 4.24	(12) 18.01 ± 7.51
(iii)	(00) 0.06 ± 0.03	(00) 2.01 ± 0.10	(00) 22.41 ± 1.21	(00) 3.94 ± 0.25	(00) 0.9892 ± 0.0014	(00) 0.7532 ± 0.0366	(00) 0.6407 ± 0.0242	(00) 21.17 ± 0.48	(00) 58.88 ± 1.10	(00) 35.20 ± 0.77
	(24) 0.37 ± 0.08	**(02) 1.97 ± 0.10**	(02) 22.04 ± 1.28	(02) 3.87 ± 0.28	(02) 0.9895 ± 0.0015	(02) 0.7685 ± 0.0368	(02) 0.6632 ± 0.0267	(02) 21.31 ± 0.52	(02) 50.83 ± 2.38	(01) 33.74 ± 0.93
	(10) 1.07 ± 0.17	(08) 2.02 ± 0.11	**(08) 18.29 ± 0.88**	**(08) 3.21 ± 0.18**	**(08) 0.9928 ± 0.0008**	**(08) 0.8241 ± 0.0278**	**(08) 0.7267 ± 0.0176**	**(08) 22.93 ± 0.43**	**(13) 22.46 ± 2.64**	**(17) 14.93 ± 1.75**
	(12) 17.81 ± 3.65	(07) 2.01 ± 0.11	(26) 21.18 ± 2.47	(07) 3.71 ± 0.24	(07) 0.9902 ± 0.0012	(07) 0.7900 ± 0.0335	(07) 0.6850 ± 0.0228	(26) 21.69 ± 1.01	(12) 27.14 ± 1.00	(12) 19.55 ± 2.73
(iv)	(00) 0.05 ± 0.01	(00) 3.37 ± 0.78	(00) 19.81 ± 1.86	(00) 5.88 ± 0.97	(00) 0.9777 ± 0.0065	(00) 0.6818 ± 0.0364	(00) 0.5928 ± 0.0236	(00) 22.26 ± 0.83	(00) 63.62 ± 1.17	(00) 38.89 ± 1.78
	(24) 0.35 ± 0.12	**(02) 3.30 ± 0.76**	(02) 19.73 ± 1.85	(02) 5.85 ± 0.97	(02) 0.9779 ± 0.0064	(02) 0.6936 ± 0.0360	(02) 0.6045 ± 0.0235	(02) 22.29 ± 0.83	(02) 57.02 ± 1.91	(23) 37.30 ± 1.19
	(15) 0.87 ± 0.13	(08) 3.37 ± 0.79	**(08) 17.04 ± 1.42**	(08) 5.08 ± 0.93	(08) 0.9834 ± 0.0054	**(08) 0.7598 ± 0.0344**	**(08) 0.6892 ± 0.0236**	**(08) 23.56 ± 0.73**	**(13) 32.89 ± 2.81**	**(17) 16.15 ± 0.76**
	(12) 14.28 ± 2.60	(07) 3.37 ± 0.78	(26) 19.24 ± 3.99	**(26) 4.45 ± 2.14**	**(26) 0.9847 ± 0.0125**	(07) 0.7236 ± 0.0333	(07) 0.6419 ± 0.0227	(26) 22.64 ± 1.88	(12) 35.74 ± 0.94	(12) 19.67 ± 0.86

6. Conclusions

A practical overview and comparison analysis has been presented on the performance of single-frame super-resolution methods applied to remote sensing images. From the set of 27 methods analysed over six simulated image collections, the best four methods are **08SDS (Dong et al. 2011)**, **25FSR (Zhao et al. 2015)**, **07LSE (Freedman and R. 2011)** and **03GPP (Jian, Xu, and Shum 2008)**. These methods are of type **sparse coding**, **PSF deconvolution**, **hybrid** and **gradient profile** type, respectively. Note that experiments have been conducted using the default settings of SR codes in order to perform a comparison as fair as possible considering the high variety of techniques. However, tuning parameters of some methods might improve their corresponding results.

Although **08SDS (Dong et al. 2011)** shows the best general performance, there are several important considerations to point out. In particular three practical conclusions arise from this work: (i) learning-based methods tend to outperform the rest of SR techniques in remote sensing when considering moderate up-scaling factors and reduced levels of noise, (ii) reconstruction-based and hybrid methods are able to improve their results with higher scaling factors and levels of noise, although PSF deconvolution methods depend on the quality of the PSF kernel characterisation, and (iii) Lanczos-3 interpolation gives better results than expected especially with noisy images.

Learning-based methods use external training sets to learn high-resolution structures, therefore it seems clear that these methods are the best when relevant training examples are available. However, when not enough samples are available, reconstruction and hybrid methods are suggested instead. Within all learning techniques, sparse code methods have shown to be the most effective ones mainly because there is not any neighbourhood or functional constraint in the learning process, as in the case of neighbourhood embeddings or mappings.

In general, the increase in the up-scaling resolution factor (from 2x to 4x) and in the amount of noise (from noiseless to $\sigma = 0.05$ additive white Gaussian noise) decreases significantly the performance of all the tested SR methods. Although learning-based methods have shown to perform well under any imaging model, it is important to remark that reconstruction-based and hybrid methods tend to be more robust to high up-scaling factors and levels of noise because their performance

drops are less pronounced. As differences between LR and HR domains become bigger, it is more difficult to learn a model to effectively fill the gap between LR and HR. Considering both reconstruction-based and hybrid methods, the former type seems to be more robust to noise and the latter more robust to high scaling factors.

Finally, it is worthwhile indicating that the baseline method (**00IntL3 (Lanczos-3 Interpolation)**) provides better results than expected as compared to the most of SR methods analysed in this work.

Acknowledgment

References

A.C., Bovik. 2010. "Perceptual Video Processing: Seeing the Future." *Proceedings of the IEEE* 98 (11): 1799–1803.

Arozarena, A., Villa G., and Valcarcel N. 2005. "The National Aerial Ortophoto Program in Spain (PNOA)." In *International Cartographic Conference,* .

Baker, S., and Kanade T. 2002. "Limits on super-resolution and how to break them." *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (9): 1167–1183.

Bioucas-Dias, J., Plaza A., Camps-Valls G., Scheunders P., Nasrabadi N. M., and Chanussot J. 2013. "Hyperspectral Remote Sensing Data Analysis and Future Challenges." *IEEE Geoscience and Remote Sensing Magazine* 1 (2): 6–36.

Chavez-Roman, H., and Ponomaryov V. 2014. "Super Resolution Image Generation Using Wavelet Domain Interpolation With Edge Extraction via a Sparse Representation." *IEEE Geoscience and Remote Sensing Letters* 11 (10): 1777–1781.

Dai, D., Timofte R., and Van Gool L. 2015. "Jointly Optimized Regressors for Image Super-resolution." In *Eurographics,* .

Demirel, H., and Anbarjafari G. 2011. "Discrete Wavelet Transform-Based Satellite Image Resolution Enhancement." *IEEE Transactions on Geoscience and Remote Sensing* 49 (6): 1997–2004.

Dong, C., Loy C., He K., and Tang X. 2014. "Learning a Deep Convolutional Network for Image Super-Resolution." In *ECCV,* Vol. 8692184–199.

Dong, C., Loy C., He K., and Tang X. 2015. "Image Super-Resolution Using Deep Convolutional Networks." *IEEE Trans. Pattern Anal. Mach. Intell.* .

Dong, W., Zhang D., Shi G., and Wu X. 2011. "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization." *Image Processing, IEEE Transactions on* 20 (7): 1838–1857.

Freedman, G., and Fattal R. 2011. "Image and Video Upscaling from Local Self-examples." *ACM Transactions on Graphics* 30 (2): 1–11.

Glasner, D., Bagon S., and Irani M. 2009. "Super-Resolution from a Single Image." In *International COnference on Computer Vision,* 349–356.

Gonzalez, Rafael C., and Richard E. Woods. 2006. *Digital Image Processing (3rd Edition).* Prentice-Hall, Inc.

H., Chang, Yeung D. T., and Xiong Y. 2004. "Super-resolution through neighbor embedding." In *IEEE Computer Vision and Pattern Recognition,* .

He, H., and Siu W.-C. 2011. "Single image super-resolution using Gaussian process regression.." In *IEEE Computer Vision and Pattern Recognition,* 449–456.

Huang, J.B., Singh A., and Ahuja N. 2015. "Single Image Super-Resolution From Transformed Self-Exemplars." In *IEEE Conference on Computer Vision and Pattern Recognition,* 5197–5206.

Irani, M., and Peleg S. 1991. "Improving Resolution by Image Registration." *CVGIP: Graph. Models Image Process.* 53 (3): 231–239.

Jian, S., Z. Xu, and H.-Y. Shum. 2008. "Image super-resolution using gradient profile prior.." In *CVPR,* .

Kim, K. I., and Kwon Y. 2010. "Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior." *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (6): 1127–1133.

Li, L., Y. Chen, Xu T., Liu R., and Huang C. 2015a. "Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization." *ISPRS Journal of Photogrammetry and Remote Sensing* 101: 10–21.

Li, L., Y. Chen, Xu T., Liu R., Shi K., and Huang C. 2015b. "Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm." *Remote Sensing of Environment* 164: 142–154.

Liu, D., Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang. 2016. "Robust Single Image Super-Resolution via Deep Networks With Sparse Prior." *IEEE Transactions on Image Processing* 25 (7): 3194–3207.

Loncan, L., L. Almeida, J. Bioucas, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, et al. 2015. "Hyperspectral pansharpening: a review." *IEEE Geoscience and Remote Sensing Magazine* 2 (3): 27–46.

Lucy, L.B. 1974. "An iterative technique for the rectification of observed distributions." *The Astronomical Journal* 79: 745.

Michaeli, T., and Irani M. 2013. "Non-parametric blind super-resolution." In *International Conference on Computer Vision (ICCV),* 945–952. IEEE.

Michaeli, T., and Irani M. 2014. "Blind Deblurring Using Internal Patch Recurrence." In *European Conference on Computer Vision,* 783–798.

Mittal, A., Moorthy A. K., and Bovik A. C. 2012. "No-reference image quality assessment in the spatial domain." *IEEE Transactions on Image Processing* 4695–4708.

Moreno, J. 2004. *SPARC Data Acquisition Report.* Technical report. https://earth.esa.int/documents/10174/21020/SPARC2004_Data_Acquisition_Report.pdf: European Space Agency.

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

Nouri, A., Charrier C., Saadane A., and Fernandez-Maloigne C. 2013. "Statistical comparison of no-reference images quality assessment algorithms." In *Colour and Visual Computing Symposium (CVCS),* 1–5.

Polatkan, G., Zhou M., Carin L., Blei D., and Daubechies I. 2015. "A Bayesian Nonparametric Approach to Image Super-Resolution." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (2): 346–358.

Purkait, P., and Chanda B. 2013. "Image Upscaling Using Multiple Dictionaries of Natural Image Patches." In *ACCV,* Vol. 7726284–295.

Ramji, R., Sakthivel Punniakodi, and P. Praveen. 2013. "Super Resolution and Image Enhancement in Remote Sensing Area." *International Journal of Engineering Research & Technology* 2 (9): 2288–2292.

Reibman, A. R., R. M. Bell, and S. Gray. 2006. "Quality assessment for super-resolution image enhancement." In *International Conference on Image Processing,* 2017–2020.

Shan, Q., Jia J., and A. Agarwala. 2008. "High-quality Motion Deblurring from a Single Image." *ACM Transactions on Graphics (SIGGRAPH)* .

Shan, Q., Li Z., Jia J., and Tang C. K. 2008. "Fast Image/Video Upsampling." *ACM Transactions on Graphics (SIGGRAPH ASIA)* .

Simoes, M., Bioucas-Dias J., Almeida L. B., and Chanussot J. 2015. "A Convex Formulation for Hyperspectral Image Superresolution via Subspace-Based Regularization." *IEEE Transactions on Geoscience and Remote Sensing* 53 (6): 3373–3388.

Suganya, P., N. Mohanapriya, and A. Vanitha. 2013. "Survey on Image Resolution Tech-

niques for Satellite Images." *International Journal of Computer Science and Information Technologies* 4 (6): 835–838.

Timofte, R., De Smet V., and Van Gool L. 2013. "Anchored Neighborhood Regression for Fast Example-Based Super-Resolution." In *IEEE International Conference on Computer Vision,* 1920–1927.

Turkowski, K. 1990. "Graphics Gems." chap. Filters for Common Resampling Tasks, 147–165. Academic Press Professional, Inc.

Veganzones, M., Simoes M., Licciardi G., Yokoya N., Bioucas-Dias J., and Chanussot J. 2016. "Hyperspectral super-resolution of locally low rank images from complementary multisource data." *IEEE Transactions on Image Processing* 25 (1): 274–288.

Vivone, Gemine, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio Licciardi, R. Restaino, and Lucien Wald. 2015. "A critical comparison of pansharpening algorithms." *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* 53 (5): 2565–2586.

Wald, L., Ranchin T., and Mangolini M. 1997. "Fusion of satellite images of different spatial resolution: Assessing the quality of resulting images." *Photogrammetric Engineering and Remote Sensing* 691–699.

Wang, Z., Bovik A. C., Sheikh H. R., and Simoncelli E. P. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612.

Yang, C.-H., Ma C., and Yang M.-H. 2014. "Single-Image Super-Resolution: A Benchmark." In *European Conference on Computer Vision,* .

Yang, C.-Y., and Yang M.-H. 2013. "Fast Direct Super-Resolution by Simple Functions." In *International Conference on Computer Vision,* .

Yang, Chih-Yuan, Chao Ma, and Ming-Hsuan Yang. 2014. "Single-Image Super-Resolution: A Benchmark." In *European Conference on Computer Vision,* .

Yang, D., Z. Li, Y. Xia, and Z. Chen. 2015. "Remote sensing image super-resolution: Challenges and approaches." In *International Conference on Digital Signal Processing,* 196–200.

Yang, J., Wright J., Huang T. S., and Ma Y. 2010. "Image Super-resolution via Sparse Representation." *IEEE Transactions on Image Processing* 19 (11): 2861–2873.

Yang, J., Wright J., Ma Y., and Huang T. 2008. "Image superresolution as sparse representation of raw image patches." In *IEEE Computer Vision and Pattern Recognition,* .

Yuhas, R. H., Goetz A. F. H., and Boardman J. W. 1992. "Discrimination among semi-arid landscape endmembers using the Spectral AngleMapper (SAM) algorithm." In *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop,* 147–149.

Zeyde, R., Elad M., and Protter M. 2012. "On Single Image Scale-up Using Sparse-representations." In *International Conference on Curves and Surfaces,* 711–730.

Zhang, Hongyan, Zeyu Yang, Liangpei Zhang, and Huanfeng Shen. 2014. "Super-Resolution Reconstruction for Multi-Angle Remote Sensing Images Considering Resolution Differences." *Remote Sensing* 6: 637–657.

Zhao, N., Wei Q., Basarab A., Kouame D., and J.-Y. Tourneret. 2015. "Fast Single Image Super-Resolution." *CoRR* abs/1510.00143.