To appear in *Remote Sensing Letters*
Vol. 00, No. 00, Month 20XX, 1–10

RESEARCH LETTER

Latent topic-based super-resolution for remote sensing

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

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

(*December 2016*)

This letter presents a novel single-image super-resolution approach based on latent topics specially designed to remote sensing imagery. The proposed approach pursues to super-resolve topics uncovered from low-resolution images instead of super-resolving low-resolution patches themselves. An experimental comparison is conducted using nine different super-resolution methods over four aerial image datasets. Experiments revealed the potential of topic models in remote sensing super-resolution by reporting that the proposed approach is able to provide a competitive advantage especially in low noise conditions.

Keywords: super-resolution; latent topics; LDA; image quality assessment;

1. Introduction

Single-image Super-Resolution (SR) is aimed at improving image resolution beyond the acquisition sensor limits using for that purpose a single image of the objective scene. This kind of image processing technology is especially attractive to remote sensing in satellite and airborne missions that use relatively inexpensive sensors or have a long revising time to obtain multiple consistent observations of the same point on earth. In these scenarios, single-image SR provides the opportunity to offer new super-resolved data products in order to cope with the increasing demand of remote sensing-related applications and challenges (Bioucas-Dias et al. 2013).

Broadly speaking, single-image SR algorithms can be categorized into two different groups (Nasrollahi and Moeslund 2014), image REconstruction (RE) and image LEarning (LE) methods. While RE methods do not require any kind of training process, LE techniques are able to obtain better results by learning the relationships between Low-Resolution (LR) and High-Resolution (HR) image details from an external image training set. Even though LE approaches have shown to be effective under a wide range of conditions, each learning model has its own limitations and therefore SR performance highly depends on the application field.

Recent research lines try to overcome current LE limitations by taking advantage from a visual interpretation point of view of the so-called *image semantics* (Timofte, Smet, and Gool 2016), that is, modelling the image visual interpretation. The rationale behind this methodology is based on learning a specific model for each semantic concept appearing in the training set and then super-resolving each input patch using the most suitable model. Typically, semantic concepts are defined by

*Corresponding author. Email: rufernan@uji.es

an initial clustering process and a similarity-based classifier is used to predict the input image semantic concepts.

In contrast to standard images, remote sensing imagery have a more complex nature because they are usually fully-focused multi-band shots with plenty of different textured details within the same image. The high intricacy of satellite and aerial imagery makes the classical classification-based approach unable to capture complex visual concepts and relationships. This fact eventually limits the SR semantic power in remote sensing (Ramji, Punniakodi, and Praveen 2013).

The main objective of this work is to improve single-image SR in remote sensing by enhancing its semantic level through latent topics (Blei 2012). These kinds of statistical models are able to uncover the hidden patterns of a document collection and thus they have been successfully used in many other related image processing applications to provide data with a higher level of semantic understanding. Due to the special relevance of semantics in remote sensing, latent topics may be an useful tool when super-resolving aerial imagery. The proposed Topic-based Super-Resolution approach (TSR) is based on super-resolving image latent topics instead of image patches themselves in order to manage the semantics variability through the patterns defined by topics.

The rest of the paper is organized as follows. Section 2 proposes the topic-based SR framework specially designed to remote sensing imagery. Section 3 presents the experimental part of the work where eight SR methods are compared against the proposed approach using four remote sensing datasets. Finally, Section 4 discusses the obtained results and conclusions are given in Section 5.

2. Super-resolution framework based on latent topics

In order to super-resolve multi-spectral remote sensing images, we follow the standard SR procedure based on the $\mathbf{YC_bC_r}$ color space transformation (Nasrollahi and Moeslund 2014). Initially, input **RGB** bands are converted to the $\mathbf{YC_bC_r}$ color space. Then, the luminance channel $\mathbf{Y}$ is super-resolved and the rest of the components, i.e. $\mathbf{C_b}$ (blue difference chroma), $\mathbf{C_r}$ (red difference chroma) and any other remainder spectral band, are interpolated to the target resolution. Finally, the inverse $\mathbf{YC_bC_r}$ transformation is used to generate the super-resolved output.

Regarding the framework image characterisation, we make use of the Bag-of-Words (BoW) approach (Zhang, Jin, and Zhou 2010) adapted to the image domain in order to enable the use of topic models over images. Specifically, vectorised image patches are considered topic model documents (d), pixel positions within patches define the vocabulary words of the collection (w) and document word-counts are represented by pixel luminance values. Note that considering an image size of $(r \times c)$, a patch size of $(s \times s)$, where $s = 2x + 1$, and one pixel of patch overlapping, this characterisation generates a total of $D = (r - 2x)(c - 2x)$ documents with a $W = s^2$ vocabulary size.

Figure 1 shows the four stages of the proposed Topic-based Super-Resolution (TSR) approach: (*1*) topic super-resolution learning (Section 2.1), (*2*) topic-document estimation (Section 2.2), (*3*) topic-based image reconstruction (Section 2.3) and (*4*) post-processing (Section 2.4). Note that stage (*1*) corresponds to the training step (computed off-line) and stages from (*2*) to (*4*) are the test step (carried out under demand).

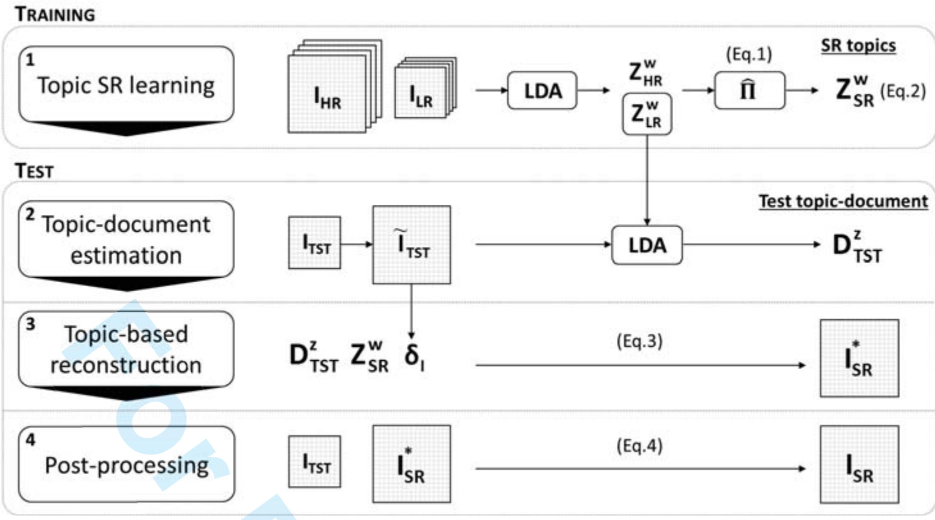


Figure 1.: The Topic-based Super-Resolution (TSR) approach graphical description.

2.1. *Topic super-resolution learning*

As a learning SR method, the proposed approach (TSR) requires several training examples to learn the relationships between the LR and the HR domains. Specifically, this stage aims at: (i) extracting HR training topics, (ii) extracting LR training topics and (iii) learning the connection between LR and HR topics. First, HR training images $\mathbf{I}_{\mathrm{HR}}$ are characterised according to the aforementioned BoW approach as $\mathbf{D}^{\mathrm{w}}_{\mathrm{HR}} = \{\boldsymbol{p}(w_i|d_n)\}$ where $i \in [1, W]$ and $n \in [1, D]$. Then, Latent Dirichlet Allocation (LDA) model (Blei, Ng, and Jordan 2003) is used to uncover K topics represented by conditional probabilities $\mathbf{Z}^{\mathrm{w}}_{\mathrm{HR}} = \{\boldsymbol{p}(w_i|z_j)\}$, where $j \in [1, K]$. Second, LR training images $\mathbf{I}_{\mathrm{LR}}$ are up-scaled to the HR size by a bi-cubic interpolation as $\widetilde{\mathbf{I}}_{\mathrm{LR}}$. Then, LR topics $\mathbf{Z}^{\mathrm{w}}_{\mathrm{LR}}$ and topic-document descriptions $\mathbf{D}^{\mathrm{z}}_{\mathrm{LR}} = \{\boldsymbol{p}(z_j|d_n)\}$ are uncovered from $\widetilde{\mathbf{I}}_{\mathrm{LR}}$ following the same procedure described for HR training images. Third, Eq. (1) is used to obtain the optimal permutation $\hat{\mathbf{\Pi}}$ between LR and HR topics that minimises the ℓ^2-norm reconstruction error over HR training documents. Finally, Eq. (2) generates the super-resolved topics $\mathbf{Z}^{\mathrm{w}}_{\mathrm{SR}}$.

$$\hat{\mathbf{\Pi}} = \underset{\mathbf{\Pi}^*}{\arg\min} \parallel \mathbf{D}^{\mathrm{w}}_{\mathrm{HR}} - \mathbf{D}^{\mathrm{z}}_{\mathrm{LR}}\, \mathbf{\Pi}^*\, \mathbf{Z}^{\mathrm{w}}_{\mathrm{HR}} \parallel_2 \tag{1}$$

$$\mathbf{Z}^{\mathrm{w}}_{\mathrm{SR}} = \hat{\mathbf{\Pi}}\, \mathbf{Z}^{\mathrm{w}}_{\mathrm{HR}} \tag{2}$$

In order to alleviate Eq. (1) computational cost, we reduce the solution space adding a constraint on the shape of $\mathbf{\Pi}^*$. In particular, we use the Pearson Correlation Coefficient between $\mathbf{Z}^{\mathrm{w}}_{\mathrm{LR}}$ and $\mathbf{Z}^{\mathrm{w}}_{\mathrm{HR}}$ to explore only those permutations that replace LR topics by any of the three most correlated topics in the HR topic domain.

2.2. *Topic-document estimation*

This stage deals with estimating the LR input test image $\mathbf{I}_{\mathrm{TST}}$ representation in the LR topic space $\mathbf{Z}^{\mathrm{w}}_{\mathrm{LR}}$ learnt in the training stage. Initially, $\mathbf{I}_{\mathrm{TST}}$ is up-sampled to the target resolution using a bi-cubic interpolation. Then, this interpolated image $\widetilde{\mathbf{I}}_{\mathrm{TST}}$

is characterised following the BoW approach as $\mathbf{D}^{\mathrm{w}}_{\mathrm{TST}} = \{\boldsymbol{p}(w_i|d_n)\}$. Finally, LDA (Blei, Ng, and Jordan 2003) is used to estimate the topic-document descriptions $\mathbf{D}^{\mathrm{z}}_{\mathrm{TST}} = \{\boldsymbol{p}(z_j|d_n)\}$ by fixing the set of topics to the LR training topics $\mathbf{Z}^{\mathrm{w}}_{\mathrm{LR}}$.

2.3. *Topic-based image reconstruction*

The objective of this stage is to generate the super-resolved reconstruction result $\mathbf{I}^*_{\mathrm{SR}}$ from both $\mathbf{D}^{\mathrm{z}}_{\mathrm{TST}}$ and $\mathbf{Z}^{\mathrm{w}}_{\mathrm{SR}}$ distributions using Eq. (3). First, the document-word distribution $\mathbf{D}^{\mathrm{w}}_{\mathrm{SR}} = \{\boldsymbol{p}(w_i|d_n)\}$ is estimated. Then, pixel luminance values are recovered multiplying probabilities $\mathbf{D}^{\mathrm{w}}_{\mathrm{SR}}$ by the prior $\boldsymbol{\delta}_{\mathrm{I}}$ which contains $\widetilde{\mathbf{I}}_{\mathrm{TST}}$ document word-counts. Finally, the super-resolved image $\mathbf{I}^*_{\mathrm{SR}}$ is rebuilt by the operator $\mathcal{W}$ which averages document-word contributions to the final image pixel positions. For this operator, we adopt a Gaussian-like windowing function (Alliney and Morandi 1986) in order to alleviate some possible misregistration effects when reconstructing the super-resolved image from documents.

$$\mathbf{I}^*_{\mathrm{SR}} = \mathcal{W}\left(\overbrace{(\mathbf{D}^{\mathrm{z}}_{\mathrm{TST}}\ \mathbf{Z}^{\mathrm{w}}_{\mathrm{SR}})}^{\mathbf{D}^{\mathrm{w}}_{\mathrm{SR}}}\ \overbrace{\boldsymbol{\delta}_{\mathrm{I}}}^{\texttt{prior}} \right) \tag{3}$$

2.4. *Post-processing*

The final stage of the proposed approach is a post-processing step (Timofte, Rothe, and Gool 2016) to enforce a global reconstruction constraint over the output result. The aim is to mitigate possible deviations between the LR observation $\mathbf{I}_{\mathrm{TST}}$ and the final super-resolved image $\mathbf{I}_{\mathrm{SR}}$. Eq. (4) illustrates the process.

$$\mathbf{I}_{\mathrm{SR}} = \arg\min_{\mathbf{I}_{\mathrm{SR}}} \ \| \,\mathcal{D}\,\mathcal{B}\,\mathbf{I}_{\mathrm{SR}} - \mathbf{I}_{\mathrm{LR}} \,\|_2 \ + \ \alpha \,\|\, \mathbf{I}_{\mathrm{SR}} - \mathbf{I}^*_{\mathrm{SR}} \,\|_2 \tag{4}$$

In this expression, $\mathcal{D}$ and $\mathcal{B}$ represent the decimating and blurring operators respectively, $\mathbf{I}^*_{\mathrm{SR}}$ is the initial guess of the solution provided by Eq. (3) and $\mathbf{I}_{\mathrm{SR}}$ is the final output after the optimisation process which is performed by gradient descent. Note that this process aims at balancing both the fitting of the final output image with the initial LR input, on the one hand and the fitting of the solution with itself by a factor α, on the other hand.

3. Experiments

3.1. *Datasets*

The remote sensing images used in this work have been selected from the open-access orthoimages of the Spanish National Aerial Ortophoto Program (PNOA) (Arozarena, G., and N. 2005). These RGB images are available on the Spanish National Geographic Institute (IGN) website (ign 2016) and they have a resolution of 0.25 mpp (meters per pixel).

A total of eight 512×512 images (Fig. 2) have been extracted from the Alicante

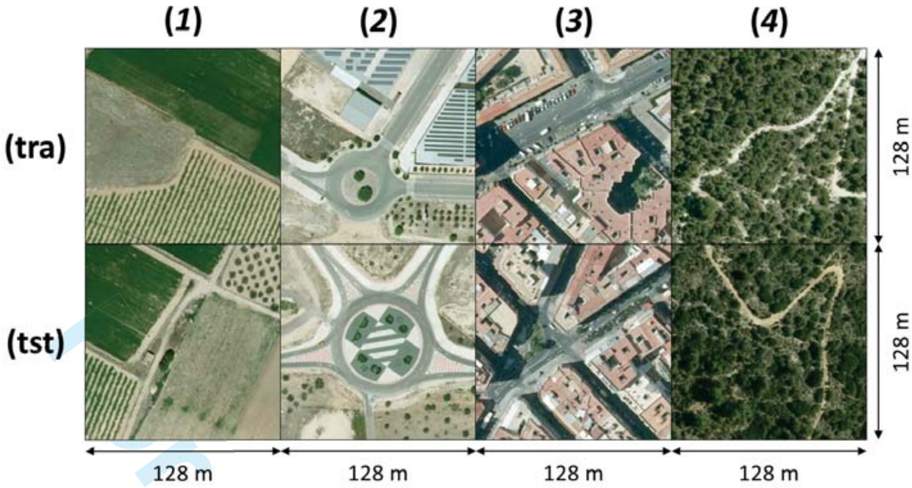


Figure 2.: Training (tra) and test (tst) HR images (RGB, 512×512, 0.25mpp) belonging to four different scenarios, from (*1*) to (*4*).

Table 1.: The four LR datasets considered for the experiments. Gaussian blur (G) and Aditive White Gaussian Noise (AWGN) for blurring ($\mathcal{B}$) and noise operators ($\mathcal{N}$).

Dataset	**Blurring** ($\mathcal{B}$)	**Decimation** ($\mathcal{D}$)	**Noise** ($\mathcal{N}$)	$\mathbf{I}_{\mathrm{LR}}$ **size**
LR2xbd	$\mathrm{G}(\mu = 0, \sigma = 1)$	$2\times$	-	256×256 (0.5 mpp)
LR4xbd	$\mathrm{G}(\mu = 0, \sigma = 1)$	$4\times$	-	128×128 (1.0 mpp)
LR2xbdn05	$\mathrm{G}(\mu = 0, \sigma = 1)$	$2\times$	$\mathrm{AWGN}(\sigma = 0.05)$	256×256 (0.5 mpp)
LR4xbdn05	$\mathrm{G}(\mu = 0, \sigma = 1)$	$4\times$	$\mathrm{AWGN}(\sigma = 0.05)$	128×128 (1.0 mpp)

area (Valencian Community) belonging to four different scenarios: (*1*) agricultural, (*2*) industrial, (*3*) urban and (*4*) forest type scenes. These high-resolution images have been used to generate four datasets (Table 1) by considering four different imaging models with the form $\mathbf{I}_{\mathrm{LR}} = \mathcal{D}(\mathcal{B}(\mathbf{I}_{\mathrm{HR}})) + \mathcal{N}$, where $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{N}$ represent the blurring, decimation and noise operators, respectively.

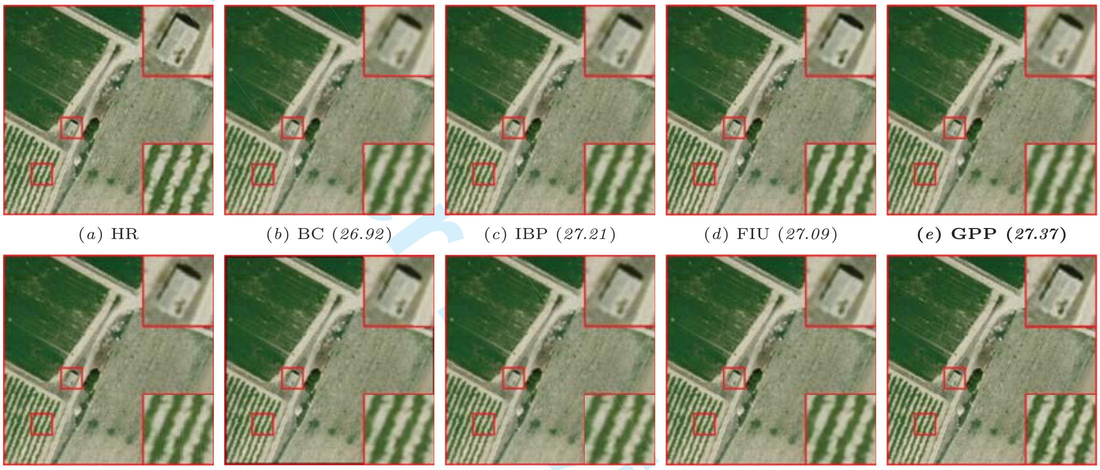
3.2. *Experimental setting*

This section describes the experimental protocol used in this work. The proposed TSR approach has been tested together with seven different SR algorithms and the bi-cubic interpolation (baseline method) over the four aforementioned LR datasets in order to generate 512×512 super-resolved images. Table 2 shows a brief summary of the methods.

All tested methods have been used considering their corresponding default settings because they are supposed to provide the most general scheme to super-resolve the high variety of remote sensing images we include in this work, that is, agricultural, industrial, urban and forest images. Learning methods have been trained using (tra) HR images of Fig. 2 and the corresponding (tra) LR images for each dataset. In the case of three of them (VSR, ANR and BSR), the number of dictionary atoms has been fixed to 1000. Regarding the proposed TSR method, we have considered a patch size of 17×17, a number of topics (K) equal to 1000 and a post-processing step using a Gaussian blurring operator with $\sigma = 0.6$ and 100 back-projection iterations.

Table 2.: Methods considered for the experiments. Further details can be found in the corresponding references.

SR method	Type	Description	Reference
BCI	Baseline	Bi-cubic interpolation	(Nasrollahi and Moeslund 2014)
IBP	Reconstruction	Iterative back projection	(Irani and Peleg 1991)
FIU	Reconstruction	Deconvolution	(Shan et al. 2008)
GPP	Reconstruction	Gradient profile	(Sun, Xu, and Shum 2008)
VSR	Learning	Sparse coding	(Yang et al. 2010)
ANR	Learning	Neighbourhood embedding	(Timofte, De Smet, and Van Gool 2013)
BSR	Learning	Bayesian mapping	(Polatkan et al. 2015)
SRI	Hybrid (RE/LE)	Scale patch redundancy	(Glasner, Bagon, and Irani 2009)
TSR	Learning	Latent topic-based	**(Proposed approach)**



(*a*) HR (*b*) BC (*26.92*) (*c*) IBP (*27.21*) (*d*) FIU (*27.09*) **(*e*) GPP (*27.37*)**
(*f*) VSR (*27.24*) (*g*) ANR (*27.01*) (*h*) BSR (*24.59*) (*i*) SRI (*27.02*) (*j*) TSR (*27.36*)

Figure 3.: SR results for the test image (*1*) of **LR2xbd** dataset. For each result, PSNR (dB) values in brackets. The best PSNR value is highlighted in bold.

Three different image quality metrics (Nasrollahi and Moeslund (2014)) are used in this work, PSNR (Peak Signal to Noise Ratio), SSIM (Structural SIMilarity) index and NIQE (Natural Image Quality Evaluator). Note that PSNR and SSIM use a reference HR image whereas NIQE is a metric without reference. The higher the PSNR and SSIM values, the better the image quality and the higher the NIQE value the worse the image quality. It should be noted that an 8-pixel border has been discarded to compute image quality metrics because some of the tested methods do not super-resolve image borders.

3.3. *Results*

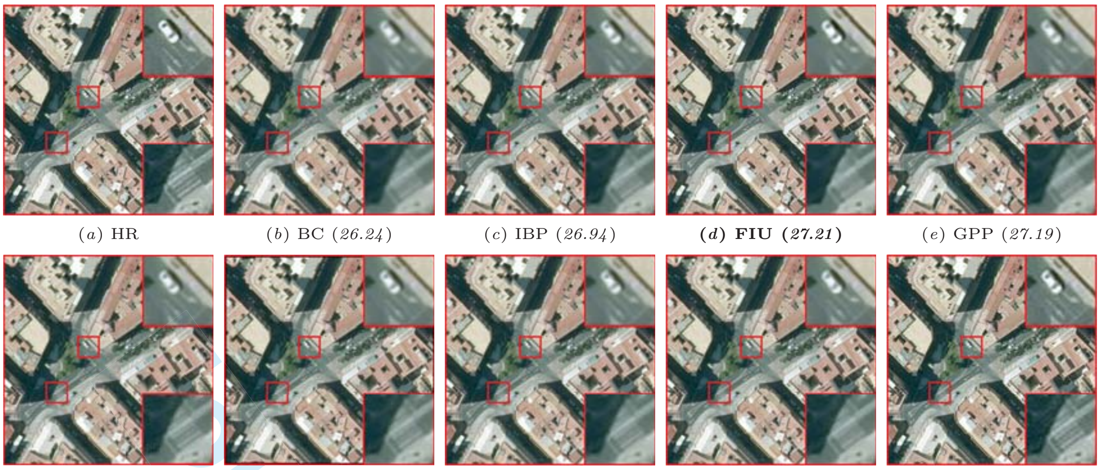
Tables 3 and 4 present the results of super-resolving (tst) LR images of Fig. 2 to achieve a final resolution of 512×512. SR methods are shown in columns while datasets, test images and quality metrics in rows. In addition, some visual results on database **LR2xbd** are shown in Figures 3 and 4 in order to provide a qualitative super-resolution assessment.

Table 3.: Super-Resolution assessment for noiseless datasets **LR2xbd** and **LR4xbd**. In rows, super-resolved test images, from *(1)* to *(4)*, and metrics PSNR (dB), SSIM and NIQE. Tested SR methods appear in columns. The best result for each image and metric is highlighted in bold.

Database	Test image	Metric	SR method code								
			BCI	**IBP**	**FIU**	**GPP**	**VSR**	**ANR**	**BSR**	**SRI**	**TSR**
LR2xbd	(1)	PSNR	*26.92*	*27.21*	*27.09*	***27.37***	*27.24*	*27.01*	*24.59*	*27.02*	*27.36*
		SSIM	*0.86*	***0.88***	*0.85*	*0.87*	*0.88*	*0.88*	*0.86*	*0.87*	*0.86*
		NIQE	*18.30*	*15.48*	*22.34*	*20.50*	*15.41*	*11.14*	*14.97*	*15.72*	***4.70***
	(2)	PSNR	*29.52*	*30.06*	*30.05*	***30.33***	*30.18*	*29.70*	*24.37*	*29.80*	*30.06*
		SSIM	*0.93*	*0.94*	*0.93*	***0.94***	*0.94*	*0.94*	*0.92*	*0.94*	*0.94*
		NIQE	*19.09*	*15.87*	*20.90*	*20.56*	*16.12*	*11.88*	*13.05*	*15.94*	***3.16***
	(3)	PSNR	*26.24*	*26.94*	***27.21***	*27.19*	*27.11*	*26.57*	*24.17*	*26.60*	*27.05*
		SSIM	*0.92*	*0.94*	*0.93*	*0.93*	*0.94*	*0.93*	*0.92*	*0.93*	***0.94***
		NIQE	*18.57*	*16.75*	*20.10*	*19.43*	*16.40*	*14.13*	*16.10*	*16.04*	***4.58***
	(4)	PSNR	*25.74*	*26.26*	*26.15*	***26.48***	*26.32*	*26.15*	*25.52*	*26.05*	*26.39*
		SSIM	*0.89*	*0.91*	*0.90*	*0.91*	*0.91*	*0.91*	*0.89*	*0.90*	***0.91***
		NIQE	*18.04*	*15.74*	*20.08*	*20.23*	*16.61*	*9.52*	*13.58*	*15.47*	***6.35***
LR4xbd	(1)	PSNR	*23.30*	*23.35*	***23.43***	*23.30*	*23.32*	*23.29*	*21.12*	*23.15*	*23.39*
		SSIM	*0.64*	*0.65*	*0.64*	*0.64*	*0.66*	*0.65*	*0.63*	*0.62*	***0.66***
		NIQE	*37.49*	*35.60*	*33.65*	*34.74*	*29.44*	*28.91*	*33.10*	*22.04*	***16.49***
	(2)	PSNR	*24.79*	*24.85*	***25.02***	*24.80*	*24.86*	*24.72*	*20.70*	*24.58*	*24.95*
		SSIM	*0.73*	*0.74*	*0.74*	*0.73*	*0.75*	*0.73*	*0.72*	*0.71*	***0.75***
		NIQE	*41.25*	*41.53*	*42.52*	*41.97*	*29.44*	*30.07*	*34.26*	*30.11*	***19.95***
	(3)	PSNR	*21.21*	*21.31*	***21.51***	*21.23*	*21.40*	*21.21*	*19.65*	*20.98*	*21.34*
		SSIM	*0.67*	*0.68*	*0.69*	*0.67*	*0.70*	*0.68*	*0.67*	*0.64*	***0.70***
		NIQE	*36.70*	*33.15*	*32.24*	*42.18*	*32.08*	*26.76*	*28.03*	*22.74*	***14.27***
	(4)	PSNR	*22.09*	*22.15*	*22.21*	*22.09*	*22.20*	*22.12*	*21.64*	*21.94*	***22.25***
		SSIM	*0.59*	*0.60*	*0.60*	*0.59*	*0.63*	*0.61*	*0.60*	*0.56*	***0.63***
		NIQE	*45.56*	*47.64*	*41.42*	*45.86*	*35.60*	*25.79*	*36.27*	*30.05*	***21.46***

Table 4.: Super-Resolution assessment for noisy datasets **LR2xbdn05** and **LR4xbdn05**. In rows, super-resolved test images, from *(1)* to *(4)*, and metrics PSNR (dB), SSIM and NIQE. Tested SR methods appear in columns. The best result for each image and metric is highlighted in bold.

Database	Test image	Metric	SR method code								
			BCI	**IBP**	**FIU**	**GPP**	**VSR**	**ANR**	**BSR**	**SRI**	**TSR**
LR2xbdn05	(1)	PSNR	*26.26*	*26.16*	*26.38*	***26.46***	*26.20*	*25.88*	*23.99*	*26.11*	*26.32*
		SSIM	*0.79*	*0.79*	*0.79*	***0.80***	*0.79*	*0.78*	*0.77*	*0.79*	*0.80*
		NIQE	*14.10*	*12.07*	*20.39*	*16.59*	*14.11*	*9.13*	*12.69*	*12.61*	***5.23***
	(2)	PSNR	*28.38*	*28.22*	***28.71***	*28.66*	*28.32*	*27.78*	*23.80*	*28.10*	*28.37*
		SSIM	*0.87*	*0.86*	***0.88***	*0.87*	*0.86*	*0.85*	*0.84*	*0.86*	*0.87*
		NIQE	*15.60*	*11.90*	*19.94*	*16.91*	*13.83*	*8.68*	*12.58*	*12.23*	***5.34***
	(3)	PSNR	*25.69*	*25.99*	***26.41***	*26.34*	*26.13*	*25.58*	*23.65*	*25.75*	*25.94*
		SSIM	*0.88*	*0.88*	***0.89***	*0.89*	*0.88*	*0.87*	*0.86*	*0.87*	*0.89*
		NIQE	*15.67*	*11.73*	*21.08*	*17.29*	*12.98*	*10.09*	*12.48*	*11.94*	***4.71***
	(4)	PSNR	*25.24*	*25.42*	*25.50*	***25.73***	*25.48*	*25.23*	*24.82*	*25.28*	*25.43*
		SSIM	*0.85*	*0.86*	***0.87***	*0.87*	*0.87*	*0.86*	*0.85*	*0.86*	*0.87*
		NIQE	*17.55*	*13.15*	*20.44*	*17.54*	*14.41*	*8.50*	*12.49*	*12.82*	***7.73***
LR4xbdn05	(1)	PSNR	*23.00*	*22.97*	***23.12***	*22.95*	*22.82*	*22.78*	*20.83*	*22.84*	*23.01*
		SSIM	*0.59*	*0.59*	***0.60***	*0.58*	*0.58*	*0.57*	*0.56*	*0.57*	*0.59*
		NIQE	*32.85*	*29.27*	*35.14*	*30.51*	*26.46*	*19.94*	*23.86*	*25.93*	***16.15***
	(2)	PSNR	*24.38*	*24.34*	***24.56***	*24.32*	*24.16*	*24.04*	*20.44*	*24.18*	*24.41*
		SSIM	*0.68*	*0.68*	***0.69***	*0.67*	*0.67*	*0.66*	*0.65*	*0.66*	*0.68*
		NIQE	*36.82*	*34.37*	*42.91*	*33.44*	*29.76*	*21.19*	*27.02*	*31.34*	***17.94***
	(3)	PSNR	*21.02*	*21.07*	***21.28***	*21.01*	*21.08*	*20.89*	*19.45*	*20.79*	*21.00*
		SSIM	*0.64*	*0.64*	***0.65***	*0.63*	*0.65*	*0.63*	*0.62*	*0.61*	*0.64*
		NIQE	*36.93*	*33.34*	*35.21*	*34.51*	*30.11*	*24.46*	*25.26*	*22.75*	***18.00***
	(4)	PSNR	*21.85*	*21.86*	***21.94***	*21.82*	*21.80*	*21.71*	*21.31*	*21.70*	*21.91*
		SSIM	*0.57*	*0.57*	*0.57*	*0.56*	*0.58*	*0.57*	*0.56*	*0.54*	***0.58***
		NIQE	*41.25*	*42.65*	*39.91*	*42.69*	*28.38*	*22.43*	*26.29*	*33.91*	***15.36***



(*a*) HR (*b*) BC (*26.24*) (*c*) IBP (*26.94*) **(*d*) FIU (*27.21*)** (*e*) GPP (*27.19*)
(*f*) VSR (*27.11*) (*g*) ANR (*26.57*) (*h*) BSR (*24.17*) (*i*) SRI (*26.60*) (*j*) TSR (*27.05*)

Figure 4.: SR results for the test image (*3*) of **LR2xbd** dataset. For each result, PSNR (dB) values in brackets. The best PSNR value is highlighted in bold.

4. Discussion

According to the quantitative evaluation reported in Tables 3 and 4, the proposed approach (TSR) is able to achieve the best result in multiple scenarios. In particular, TSR obtains the best NIQE value for all the considered datasets. Besides, PSNR and SSIM values are often within the top three results. For noiseless datasets (Table 3), TSR obtains the best average SSIM value and the second best average PSNR value. For noisy datasets (Table 4), the proposed approach reaches the second and third best average values for SSIM and PSNR metrics, respectively.

Regarding the image nature, we found that the proposed approach is specially effective over highly textured remote sensing images. In the case of agricultural (*1*) and forest (*4*) image types, TSR obtains a quantitative result very close to the best one. For industrial (*2*) and urban (*3*) images, the proposed approach has shown to be more effective under noiseless conditions.

Figures 3 and 4 show some visual results to highlight the proposed approach potential. As we can see, each SR method tends to foster a particular kind of visual features on the super-resolved output. Some methods, like IBP (Irani and Peleg 1991) or ANR (Timofte, De Smet, and Van Gool 2013), are able to obtain more defined edges, while others, like FIU (Shan et al. 2008) or SRI (Glasner, Bagon, and Irani 2009), seem more robust to noise by generating smoother super-resolved textures.

In terms of visual perceived quality, TSR is able to achieve a remarkable performance. For instance, the crop detail in Fig. 3(j) is certainly the most similar to its reference HR image in Fig. 3(a) and the car detail in Fig. 4(j) is the visually closest result to Fig. 4(a) as well. Super-resolving latent patterns instead of patches allows the proposed approach to achieve quite realistic results because LR patterns are replaced throughout the whole image by HR ones containing the same semantic meaning but more high-frequency information. Note that this kind of super-resolution is especially suitable for remote sensing due to the aerial imagery high complexity. Figure 5 shows an example of this semantic super-resolution. Comparing both arrow details, we can see how the proposed approach (TSR) changes the arrow shape but it essentially carries the same semantic visual information.

(a) HR (b) TSR

Figure 5.: Arrow detail of test image (3) of **LR2xbd** dataset.

Despite its potential, TSR has two main limitations generated by the use of the standard LDA model. First, LDA does not consider noisy document observations what makes TSR sensitive to input noise as it can be noted from Table 4. Second, standard LDA contemplates a single vocabulary, hence TSR requires an interpolation to unify both LR and HR document vocabularies. This initial interpolation introduces some aliasing errors in the super-resolved result as it can be seen in the pedestrian crossing detail of Fig. 4(j). Note that we use standard LDA for the sake of simplicity but further research may be focused on using extended models instead.

5. Conclusions and future work

In this letter, a topic-based SR framework is presented in order to show the potential of latent topics to super-resolve remote sensing images. The proposed approach takes advantage of the latent topic space semantics to super-resolve hidden patterns instead of image patches themselves. The experimental part of the work assesses the proposed approach performance over four different remote sensing datasets together with eight SR methods available in the literature. Experiments reveal that the proposed approach is able to obtain competitive results when considering a noiseless scenario.

The main conclusion that arises from the work is the importance of topic models to deal with the SR problem ill-posed nature. The acquisition process generates an information loss that makes that several super-resolved images may correspond to the same LR input. Topic-based SR tries to reduce this uncertainty by preserving the distribution of hidden patterns in the final result and this semantic connection is especially important in remote sensing because of the high complexity of aerial imagery.

In a sense, this letter encourages the use of topic models within the remote sensing SR field. Although the presented results are encouraging, more research in topic-based remote-sensing SR is required to provide a competitive approach that may be robust to both noise and aliasing. Specifically, further work is directed to extend this work in the following directions:

- An LDA-based extension to consider noisy observations when extracting LR topics in order to mitigate the resulting noise.
- A new topic model to perform the topic-based image reconstruction by managing two different vocabularies (LR and HR) in order to avoid the aliasing generated by the initial interpolation.

- Extending the proposed SR framework to a hybrid approach by exploiting the redundancy property over image scales.

Acknowledgement

This work was supported by the Spanish Ministry of Economy under the project ESP2013-48458-C4-3-P, by Generalitat Valenciana through project PROMETEO-II/2014/062, and by Universitat Jaume I through project P11B2014-09.

References

2016. "Website National Geographic Institute of Spain." `http://pnoa.ign.es/`.

Alliney, Stefano, and Carlo Morandi. 1986. "Digital Image Registration Using Projections." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (2): 222–233.

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

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.

Blei, D., A. Ng, and M. Jordan. 2003. "Latent dirichlet allocation." *Journal of Machine Learning Research* 3 (4-5): 993–1022.

Blei, David M. 2012. "Probabilistic topic models." *Comm. of the ACM* 55 (4): 77–84.

Glasner, Daniel, Shai Bagon, and Michal Irani. 2009. "Super-Resolution from a Single Image." In *IEEE Int. Conference on Computer Vision,* .

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

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

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

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

Shan, Qi, Zhaorong Li, Jiaya Jia, and Chi-Keung Tang. 2008. "Fast Image/Video Upsampling." *ACM Transactions on Graphics (SIGGRAPH ASIA)* .

Sun, Jian, Zongben Xu, and Heung-Yeung Shum. 2008. "Image super-resolution using gradient profile prior." In *IEEE Conf. on Computer Vision and Pattern Recognition,* .

Timofte, Radu, Vincent De Smet, and Luc Van Gool. 2013. "Anchored Neighborhood Regression for Fast Example-Based Super-Resolution." In *IEEE Int. Conference on Computer Vision,* 1920–1927.

Timofte, Radu, Rasmus Rothe, and Luc Van Gool. 2016. "Seven ways to improve example-based single image super resolution." In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, .

Timofte, Radu, Vincent De Smet, and Luc Van Gool. 2016. "Semantic super-resolution: When and where is it useful?." *Computer Vision and Image Understanding* 142: 1–12.

Yang, Jianchao, John Wright, Thomas S. Huang, and Yi Ma. 2010. "Image Super-resolution via Sparse Representation." *IEEE Trans. Img. Proc.* 19 (11): 2861–2873.

Zhang, Yin, Rong Jin, and Zhi-Hua Zhou. 2010. "Understanding bag-of-words model: a statistical framework." *Int. Journal of Machine Learning and Cybernetics* 1 (1): 43–52.