Sparse Multi-modal probabilistic Latent Semantic Analysis for Single-Image Super-Resolution

Ruben Fernandez-Beltran[a,*], Filiberto Pla[a]

[a] *Institute of New Imaging Technologies, Universitat Jaume I, Av. Sos Baynat s/n, 12071, Castellon de la Plana, Spain*

Abstract

This paper presents a novel single-image super-resolution (SR) approach based on latent topics in order to take advantage of the semantics pervading the topic space when super-resolving images. Image semantics has shown to be useful to relieve the ill-posed nature of the SR problem, however the most accepted clustering-based approach used to define semantic concepts limits the capability of representing complex visual relationships. The proposed approach provides a new probabilistic perspective where the SR process is performed according to the semantics encapsulated by a new topic model, the Sparse Multi-modal probabilistic Latent Semantic Analysis (sMpLSA). Firstly, the sMpLSA model is formulated. Subsequently, a new SR framework based on sMpLSA is defined. Finally, an experimental comparison is conducted using seven learning-based SR methods over three different image datasets. Experiments reveal the potential of latent topics in SR by reporting that the proposed approach is able to provide a competitive performance.

Keywords: Super-Resolution, Latent Topics, probabilistic Latent Semantic Analysis, Image Learning, Image Quality Assessment

1. Introduction

The objective of image Super-Resolution (SR) is to improve image resolution but not only by increasing the number of pixels but also by providing spatial details beyond the acquisition sensor precision. In the case of single-image SR

[*] Corresponding author. Tel.: +34 964 728 357

Email addresses: `rufernan@uji.es` (Ruben Fernandez-Beltran), `pla@uji.es` (Filiberto Pla)

(hereafter referred as SR), a single Low-Resolution (LR) image of the objective scene is used to generate the super-resolved output which pursues to recover High-Resolution (HR) features as if the input image were acquired using a sensor with a higher nominal resolution.

SR techniques have found a fertile domain in many applications where resolution enhancement is important. For instance, biometric identification, video surveillance, medical diagnosis, microscopic observation and remote sensing are some of the most popular application fields where SR is useful to overcome the acquisition sensor limits whatsoever.

1.1. Related work

In the literature, it is possible to find several quality works that provide a good overview of the existing SR algorithms [1, 2, 3, 4]. Roughly speaking, SR algorithms can be categorized into three different groups, image REconstruction (RE), image LEarning (LE) and HYbrid (HY) methods.

RE methods try to reconstruct HR details in the super-resolved output assuming a specific degradation model along the image acquisition process. The imaging model is typically defined by the concatenation of three operators, blurring, decimation and noise. As a result, RE methods can be seen as an inverse problem of deblurring, upsampling and denoising the input LR image. Each RE method makes its own assumptions to introduce a certain prior knowledge to well pose the inverse nature of the SR problem. For instance, iterative back projection [5], gradient profile prior [6] or Point Spread Function deconvolution [7, 8] are some of the most popular RE approaches. Although these and other RE methods have shown to be effective to reduce the noise as well as the blur and aliasing inherent to interpolation kernel functions, the lack of relevant high-frequency information in the LR input image limits their effectiveness to small magnification factors [9].

LE methods provide a more powerful scheme by learning the relationships between LR and HR domains from an external training set. Over the past years, different machine learning paradigms have been successfully applied in SR. Sparse coding [10], neighbourhood embedding [11] and mapping functions [12, 13] are amongst the most popular LE methods in the literature.

Sparse coding-based techniques take advantage of the fact that natural images tend to be sparse when they are characterised as a linear combination of small patches. In this way, dictionary atoms can be initially learnt by forcing

LR and HR training images to share the same sparse codes. Then, the LR input image sparse codes can be estimated using the LR dictionary and finally these sparse codes can be used over the HR dictionary to generate the final super-resolved output.

Neighbourhood embedding techniques assume that small image patches of LR images describe a low-dimensional non-linear manifold with a similar local geometry to their HR counterparts. As a result, HR patches can be generated as a weighted average of local neighbours using the same weights as those used in the LR domain. An example of this approach can be found in [11]. However, this work extends the classical idea of neighbourhood embedding by learning an initial sparse dictionary to reduce the number of atoms to perform the embedding and therefore reducing the computational time.

Mapping-based methods consider the SR task as a regression problem between the HR and LR spaces. The underlying idea is based on learning a mapping function between LR and HR images from a specific training set. Then, this function can be used to generate the final SR result from the LR input image. In the literature, we can find different kinds of techniques to perform that regression. Neural networks [12] and Bayesian models [13] are some of the most recent approaches. Despite the fact that LE methods are able to learn spatial details that are impossible to recover by RE approaches, their main limitation is based on the availability of a suitable training set containing HR images.

HY methods work towards reaching an agreement between RE and LE methods. In particular, they perform a training process but using only the LR input image. The rationale behind HY methods is based on the patch redundancy property pervading natural images which assumes that natural images tend to contain repetitive structures within the same scale and over scales as well. Taking this principle into account, it is possible to find patches which appear in a lower scale, without any blurring or decimation, and then extracting their corresponding HR counterparts from the higher scale image. Eventually, the super-resolved image can be generated using the LR/HR relations learnt across scales. Each specific HY approach defines its own assumptions about the imaging model and the patch searching criteria. For example, the work presented in [14] approximates the blur operator by a Gaussian kernel and the patch redundancy is carried out by an approximation of the nearest neighbour search. In other works, such as in [15], the blur operator is estimated at the same time as the SR output is generated through an optimisation process. Despite their advantages, HY-based methods are not able to learn as many LR/HR relations as LE methods do and this limits their potential in SR. Note that the starting point in any HY method is a LR image and the lower the resolution the lower the probability to find patches satisfying the redundancy property at a lower scale.

1.2. Current limitations and trends

LE methods have shown to be the most effective ones under a suitable training data. However, each learning model has its own generalisation constraints what makes the SR performance highly application field dependant [3]. Recent research lines try to overcome this limitation by taking advantage of the so-called *image semantics* [16], that is, modelling the image visual interpretation humans do. Uncertainty is one of the most important issues in SR because of the ill-posed nature of the problem, therefore modelling semantic concepts may help to discover semantic connections among patches and consequently to alleviate some ambiguities when super-resolving LR images. The idea behind this methodology is based on learning a specific model for each semantic concept appearing in the training data and then super-resolving the LR input image using the most suitable model for each patch.

These semantic concepts are usually defined in an unsupervised way according to an initial clustering process over training patches. Then, a classifier is trained to predict the semantic concept related to each LR input patch and therefore the corresponding SR model to be used. A representative semantic-based method can be found in [17] where authors present a SR approach that make use of the Expectation-Maximisation (EM) algorithm to initially cluster the data and then a linear regression function can be learnt for each group. Nonetheless, the high complexity of visual patterns in the image domain makes this straightforward approach unable to capture complex semantic concepts and relationships what eventually limits the semantic power in SR [16]. As a result, more research is required to keep improving the SR process via the image semantics research line.

During the last years, topic models have shown their potential to effectively cope with all kind of tasks by providing data with a higher level of semantic understanding [18]. Text categorisation [19], vocabulary reduction [20], visual encoding [21], image recognition [22] or even video retrieval [23] are some of the applications where topic models have been successfully used.

From a practical point of view, latent topics represent a kind of probabilistic models which provide methods to automatically understand and summarize data collections by means of their hidden patterns. Specifically, given the observed probability distribution $p(w|d)$, which describes a corpus of documents $D = \{d_1,d_2,...,d_M\}$ in a particular word-space $W = \{w_1,w_2,...,w_N\}$, latent topic algorithms are able to obtain two probability distributions: (1) the description of topics in words $p(w|z)$ and (2) the description of documents in topics $p(z|d)$. Within the image processing field, image patches usually represent documents, patch pixel positions in each patch generally define the vocabulary words and document word-counts are typically represented by pixel intensity values. In this scenario, latent topics can be seen as distinctive pixel distributions that represent the hidden image patterns of the input data. In other words, $p(w|z)$ is able to describe image patterns not explicitly present in the input data and consequently $p(z|d)$ characterises image patches at a higher abstraction or semantic level.

The majority of topic methods can be grouped into two model families, one based on probabilistic Latent Semantic Analysis (pLSA) [24] and another based on Latent Dirichlet Allocation (LDA) [25]. Although both pLSA and LDA models have shown to be effective in many fields [26, 27, 28, 29, 30], pLSA usually takes advantage of considering the document collection as model parameters in order to obtain a set of topics more correlated to the human judgement than the topics obtained by LDA [31].

The point which makes pLSA and other topic models a suitable tool for SR is their capability to represent samples in a higher-level characterization space, the so-called topic-space $Z = \{z_1,z_2,...,z_K\}$. In this space, documents are expressed as probability distributions according to their feature patterns instead of their low level features, which makes it easier for the documents to be managed at a higher abstraction level.

Despite the fact that several works in the literature advocate the use of topic models for semantic related image processing tasks [32, 16], there are almost no research work done within the SR field. Besides, the few works using topic models are not taking advantage of the inherent semantics of the topic-space to super-resolve images. For instance, the work presented in [33] uses pLSA just as a clustering algorithm of a LE-based approach but not as a model to super-resolve the data.

1.3. Work objectives and main contributions

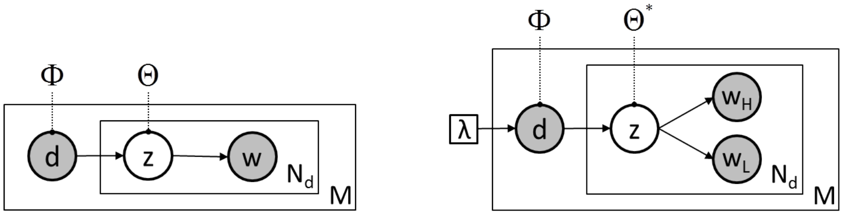
The main objective in this work is to super-resolve images following a generative framework provided by topic models in order to manage the SR semantic variability through the patterns defined by topics. That is, this work transforms the classical LE-based SR approach into a latent topic-based probabilistic approach where the SR process can be conducted according to the semantics encapsulated by the latent topic space. Specifically, we first define a pLSA-based extension, Sparse Multi-modal probabilistic Latent Semantic Analysis (sMpLSA), aimed at learning a common topic-space between LR and HR domains. Later, we use sMpLSA to super-resolve LR input images by super-resolving latent topics instead of image patches themselves. In a sense, sMpLSA allows us to tackle the SR problem as a neighbourhood embedding approach but taking into account the semantic nature of the topic space when generating the super-resolved result.

This paper extends our previous work [34] where LDA model was initially used to super-resolve remote sensing imagery. In particular, this initial approach has two main limitations. On the one hand, the use of standard LDA makes that both LR and HR topics are independent, however this is not a real premise. In fact, it seems logical to think that semantic patterns should be essentially the same whatever the resolution used to represent them. On the other hand, only remote sensing images were tested what limits the algorithm validation domain. In the present work, the SR framework is extended and the topic model is revised using more realistic assumptions which leads to an improvement of the SR performance. In addition, this work extends the experimental part with a more comprehensive experimental comparison, adding more relevant methods in the literature and using more and different application domain databases.

The rest of the paper is organized as follows: in Section 2, the proposed sMpLSA model, which is specially designed to SR, is defined. Section 3 presents the extended SR framework based on the proposed topic model. Section 4 shows the experimental part of the work where nine LE methods are tested over three different image databases considering two scaling factors. Finally, Section 5 discusses the results and Section 6 draws the main conclusions arisen from the work.

2. Sparse Multi-modal probabilistic Latent Semantic Analysis

The starting point of sMpLSA is the asymmetric formulation of pLSA (Fig. 1a), where for each document d a latent topic z is chosen conditionally to the document according to $p(z|d)$ probability distribution and then a word w is generated from that topic according to $p(w|z)$. The proposed sMpLSA model extends pLSA by considering two diverging random variables to manage different vocabulary modalities, that is, w_H to represent HR words and w_L to represent LR words. Additionally, sMpLSA incorporates a λ factor to guarantee a certain level of sparsity when representing documents in the latent topic space. Figure 1b shows the sMpLSA graphical model representation where shaded nodes represent observable random variables.



(a) pLSA

(b) sMpLSA

Figure 1: In (a), d,z,w represent the document, topic and word random variables. Φ,Θ represent the $p(z|d)$ and $p(w|z)$ model parameters. N_d,M represent the number of words in d and the total number of documents in the collection. In (b), w_H,w_L represent the HR and LR words. Finally, λ,Θ^* represent the sparsing factor and the $p(w_H|z)$,$p(w_L|z)$ parameters.

Likewise in pLSA, the sMpLSA generative process can be described as follows:

(i) A document d is chosen from $p(d)$ probability distribution.

(ii) For each one of the N_d words in the document d,

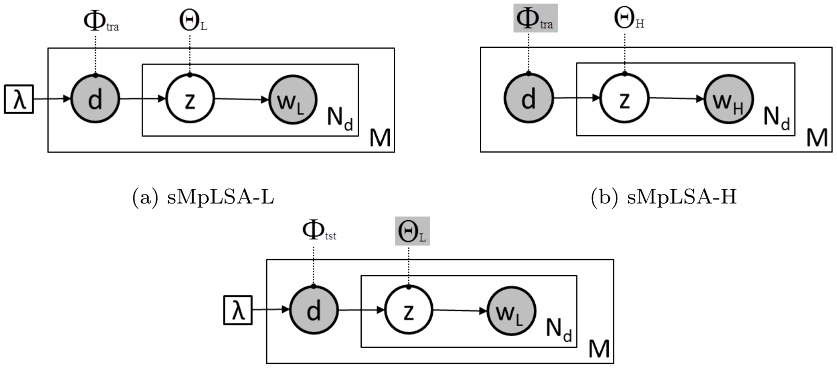
  (a) A topic z is chosen according to conditional distribution $\Phi \sim p(z|d)$ that expresses documents in topics.

  (b) Words w_H and w_L are chosen according to conditional distributions $\Theta_H \sim p(w_H|z)$ and $\Theta_L \sim p(w_L|z)$ which express topics in HR and LR words respectively. Note that we use Θ^* to refer to Θ_H and Θ_L.

2.1. Model relaxation

In order to alleviate the computational cost of managing two different vocabularies when estimating sMpLSA parameters, we propose to apply the following model relaxation based on three sequential steps:

1. **Learning LR training topics (sMpLSA-L)**: As we can see in Figure 2a, the LR part of sMpLSA corresponds to a sparse pLSA model, therefore parameters $\Phi_{tra} \sim p(z|d)$ and $\Theta_L \sim p(w_L|z)$ can be initially estimated using pLSA structure over the LR training domain.
2. **Learning HR training topics (sMpLSA-H)**: Once parameter $\Phi_{tra} \sim p(z|d)$ has been estimated, the HR part of sMpLSA model corresponds to a standard pLSA model where the $\Theta_H \sim p(w_H|z)$ parameter is the only one left to be estimated over the HR training domain. Figure 2b shows the model reduction used in the second step where shaded parameters are fixed to previously estimated values.
3. **Representing test images in LR topics (sMpLSA-tst)**: Once both training parameters $\Theta_L \sim p(w_L|z)$ and $\Theta_H \sim p(w_H|z)$ have been estimated, a sparse pLSA model can be used under demand to obtain $\Phi_{tst} \sim p(z|d_{tst})$ which represents input test documents in LR topics. Figure 2c shows the considered model reduction.



(a) sMpLSA-L

(b) sMpLSA-H

(c) sMpLSA-tst

Figure 2: sMpLSA model relaxation based on pLSA structure.

Note that this model relaxation enables dealing with the sMpLSA model with a pLSA-order computational cost.

2.2. Expectation-Maximisation learning framework

In this section, the three model reductions presented in Figure 2 are formulated. For the sMpLSA-L model, we provide a detailed description of the parameter estimation process. In the case of sMpLSA-H and sMpLSA-tst, we only provide the final expressions due to the similarity of the process.

sMpLSA-L parameters, Φ_{tra} and Θ_L, are estimated by maximising the complete log-likelihood using the Expectation-Maximisation (EM) algorithm. First, let us define the likelihood function in terms of the density function of a document collection D,

$$\mathcal{L} = p(D|\Phi_{tra},\Theta_L) = \prod_{d}^{D}\prod_{w_L}^{N_d} p(w_L,d) = \prod_{d}^{D}\prod_{w_L}^{N} p(w_L,d)^{n(w_L,d)}, \tag{1}$$

where N represents the LR vocabulary size and $n(w_L,d)$ represents the number of times the LR word w_L occurs in the document d. The joint probability $p(w_L,d)$ can be factorised according to the sMpLSA-L model as follows:

$$p(w_L,d) = \sum_{z}^{K} p(w_L|z)p(z|d)p(d) = p(d)\sum_{z}^{K} p(w_L|z)p(z|d). \tag{2}$$

Note that K represents the number of topics. Inserting Eq. (2) in Eq. (1), we obtain the expression of the complete likelihood:

$$\mathcal{L}_c = \prod_{d}^{D}\prod_{w_L}^{N}\left(p(d)\sum_{z}^{K} p(w_L|z)p(z|d)\right)^{n(w_L,d)}. \tag{3}$$

The target is to estimate the $\Phi_{tra} \sim p(z|d)$ and $\Theta_L \sim p(w_L|z)$ parameters which maximise the complete likelihood function $\mathcal{L}_c$, nonetheless multiplicative and exponential factors are hard to optimise. Due to the monotonic nature of the logarithmic function, we can equivalently maximise the complete log-likelihood (Eq. (4)) remaining the optimisation problem as Eq. (5) shows:

$$\ell_c = \log(\mathcal{L}_c) = \sum_{d}^{D}\sum_{w_L}^{N} n(w_L,d)\log\left(p(d)\sum_{z}^{K} p(w_L|z)p(z|d)\right), \tag{4}$$

$$\underset{\Phi_{tra},\Theta_L,}{\text{argmax}}(\ell_c) = \underset{\Phi_{tra},\Theta_L,}{\text{argmax}}\sum_{d}^{D}\sum_{w_L}^{N} n(w_L,d)\log\left(p(d)\sum_{z}^{K} p(w_L|z)p(z|d)\right). \tag{5}$$

Even though the performed simplifications, this expression is still hard to maximise because of the summation inside the logarithm. Taking advantage of the log function properties, we can make use of the concave version of the Jensen's Inequality as follows,

$$\begin{aligned}&\sum_{d}^{D}\sum_{w_L}^{N} n(w_L,d)\log\left(p(d)\sum_{z}^{K}p(w_L|z)p(z|d)\right)\\ \geq &\sum_{d}^{D}\sum_{w_L}^{N} n(w_L,d)p(d)\sum_{z}^{K}p(z|w_L,d)\log(p(w_L|z)p(z|d)).\end{aligned} \tag{6}$$

As a result, the expression to optimise remains as follows:

$$\mathbb{E} = \sum_{d}^{D}\sum_{w_L}^{N} n(w_L,d)p(d)\sum_{z}^{K}p(z|w_L,d)\log(p(w_L|z)p(z|d)). \tag{7}$$

Following, we introduce the normalisation constraints for parameters $p(z|d)$ and $p(w_L|z)$ by inserting the appropriate Lagrange multipliers α and β:

$$\mathbb{H}_0 = \mathbb{E} + \sum_{z}^{K}\alpha\left(1-\sum_{w}^{N}p(w|z)\right) + \sum_{d}^{D}\beta\left(1-\sum_{z}^{K}p(z|d)\right). \tag{8}$$

Finally, the solution is regularised using the sparsity factor λ to maximise the Kullback-Leibler divergence between the uniform distribution over topics (U) and the parameter $p(z|d)$:

$$\mathbb{H} = \mathbb{H}_0 + \sum_{d}^{D}\lambda(\mathrm{KL}(U|p(z|d))) = \mathbb{H}_0 - \sum_{d}^{D}\lambda\left(\frac{1}{K}\sum_{z}^{K}\log(p(z|d))\right). \tag{9}$$

To maximise the above expression we use the EM algorithm which works in two stages: (i) E-step, where given the current estimation of the parameters the expected value of the likelihood is computed (estimating the posterior probability $p(z|w_L,d)$) and (ii) M-step, where the new optimal values of the parameters are computed according to the current setting of the hidden variables.

For the M-step, we calculate Eq. (9) partial derivatives, set them equal to zero and solve the equations to estimate $p(w_L|z)$ (Eq. (10)) and $p(z|d)$ (Eq. (11)) parameters. Note that α and β multipliers can be obtained from the normalization constraint on topics and documents, respectively.

$$p(w_L|z) = \frac{\sum\limits_{d} n(w_L,d)p(d)p(z|w_L,d)}{\sum\limits_{w_L}\sum\limits_{d} n(w_L,d)p(d)p(z|w_L,d)} \tag{10}$$

$$p(z|d) = \frac{\sum\limits_{w} n(w,d)p(z|w,d)}{\sum\limits_{z}\sum\limits_{w} n(w,d)p(z|w,d)} - \frac{\lambda}{K} \tag{11}$$

For the E-step, $p(z|w_L,d)$ probabilities can be computed by applying the Bayes' rule and the chain rule as Eq. (12) shows.

$$p(z|w_L,d) = \frac{p(w_L,d,z)}{p(w_L,d)} = \frac{p(w_L,d,z)}{\sum\limits_{z} p(w_L,d)} = \frac{p(w_L|z)p(z|d)}{\sum\limits_{z} p(w_L|z)p(z|d)} \tag{12}$$

The EM process is performed as Algorithm 1 shows. First, $p(w_L|z)$ and $p(z|d)$ are randomly initialized. Then, the E-step (Eq. (12)) and the M-step (Eqs. (10)-(11)) are alternated until $p(w_L|z)$ and $p(z|d)$ parameters converge. As convergence conditions, we use a 10^{-6} stability threshold in the difference of the log-likelihood (Eq. (4)) between two consecutive iterations or a maximum number of 1000 EM iterations.

Algorithm 1: EM algorithm for sMpLSA-L.

```
input: $n(w_L,d)$, $K$, $\lambda$
$I = 0$; $T = \infty$; $L = 0$;
$p(w_L|z)$, $p(z|d)$ random initialization;
while $(I < 1000)$ and $(T > 10^{-6})$ do
    E-step: $p(z|w_L,d) \Leftarrow$ Eq. (12);
    M-step: $p(w_L|z)$, $p(z|d) \Leftarrow$ Eqs. (10)-(11);
    $\ell_c \Leftarrow$ Eq. (4); $T = \ell_c - L$; $L = \ell_c$; $I = I + 1$;
end
```

Following the same procedure, it is possible to deduce the equations for the sMpLSA-H and sMpLSA-tst models. In particular, sMpLSA-H lacks of sparsity regularisation and the $\Phi_{tra} \sim p(z|d)$ parameter is fixed to the estimation provided by sMpLSA-L. Therefore, the M-step and E-step equations for sMpLSA-H remain as Eqs. (13)-(14) show. Note that arguments $n(w_H,d)$, K

and $p(z|d)$ are now the input of the EM process and the M-step only estimates the $\Phi_H \sim p(w_H|z)$ parameter.

$$p(w_H|z) = \frac{\sum\limits_{d} n(w_H,d)p(d)p(z|w_H,d)}{\sum\limits_{w_H}\sum\limits_{d} n(H,d)p(d)p(z|w_H,d)} \tag{13}$$

$$p(z|w_H,d) = \frac{p(w_H,d,z)}{p(w_H,d)} = \frac{p(w_H,d,z)}{\sum\limits_{z} p(w_H,d)} = \frac{p(w_H|z)p(z|d)}{\sum\limits_{z} p(w_H|z)p(z|d)} \tag{14}$$

Regarding sMpLSA-tst, this model remains essentially the same as sMpLSA-L but fixing $\Theta_L \sim p(w_L|z)$. As a result, the M-step and E-step equations are given by Eqs. (15)-(16). Besides, the EM process takes $n(w_L,d_{tst})$, K, λ and $p(w_L|z)$ as input arguments and the M-step only estimates $p(z|d_{tst})$.

$$p(z|d_{tst}) = \frac{\sum\limits_{w_L} n(w_L,d_{tst})p(z|w_L,d_{tst})}{\sum\limits_{z}\sum\limits_{w_L} n(w_L,d_{tst})p(z|w_L,d_{tst})} - \frac{\lambda}{K} \tag{15}$$

$$p(z|w_L,d) = \frac{p(w_L,d_{tst},z)}{p(w_L,d_{tst})} = \frac{p(w_L,d_{tst},z)}{\sum\limits_{z} p(w_L,d_{tst})} = \frac{p(w_L|z)p(z|d_{tst})}{\sum\limits_{z} p(w_L|z)p(z|d_{tst})} \tag{16}$$

3. SR framework based on sMpLSA

Regarding the image characterisation framework, we make use of the Bag-of-Words (BoW) approach [35] adapted to the image domain in order to enable the use of topic models over images. Specifically, vectorised image patches are considered as topic model documents, pixel positions within patches define the vocabulary words of the collection and document word-counts are represented by pixel intensity values. Note that considering an image size of $(r \times c)$, a patch size of $(s \times s)$, where $s = 2x + 1$ $\forall x \in \mathbb{N} : x > 0$, and full patch overlapping, this characterisation generates a total of $D = (r - 2x)(c - 2x)$ documents with a $W = s^2$ vocabulary size.

In order to super-resolve multi-spectral RGB images, we follow the standard SR procedure based on the $\mathrm{YC_bC_r}$ color space transformation [2]. Initially, input RGB bands are converted to the YC_bC_r color space. Then, the luminance channel Y is super-resolved and the rest of the components, i.e. C_b (black-difference) and C_r (red-difference chroma), are interpolated to the target resolution. Finally, the inverse YC_bC_r transformation is used to generate the super-resolved output.

Figure 3 shows the stages of the proposed topic-based SR framework (TSR) based on sMpLSA: (1) topic-space learning (Section 3.1), (2) document projection (Section 3.2), (3) topic-based SR (Section 3.3) and (4) post-processing (Section 3.4). Specifically, stage (1) corresponds to the training step (computed off-line) and stages from (2) to (4) are the test step (carried out under demand). Note that this framework provides a kind of modular or hierarchical approach where LR image patches are super-resolved according to the image patterns uncovered by the proposed sMpLSA model.

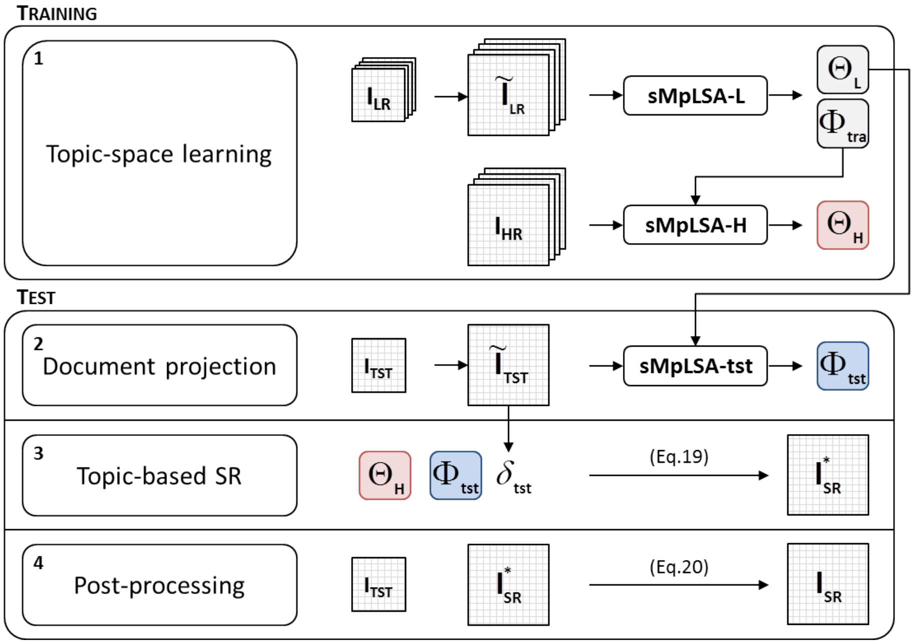


Figure 3: Graphical description of the SR framework based on sMpLSA.

3.1. Topic-space learning

As a LE method, the proposed approach requires a suitable training set in order to learn the relationships between both LR and HR image domains.

Specifically, these relationships are learned following the sMpLSA model relaxation proposed in Section 2.1. First, training LR images I_{LR} are up-sampled to the target resolution using a bi-cubic interpolation as $\tilde{I}_{LR}$ and ,subsequently, image patches are characterised as documents. Then, the sMpLSA-L model (Fig. 2a) is used to obtain the LR topics, $\Theta_L \sim p(w_L|z)$, and the shared latent topic space, $\Phi_{tra} \sim p(z|d)$, between LR and HR domains. Finally, the sMpLSA-H model (Fig. 2b) can be used to estimate the HR topics, $\Theta_H \sim p(w_H|z)$, from the HR training images I_{HR} by fixing the Φ_{tra} parameter. Note that the number of topics K and the λ sparsity factor are training parameters when applying the sMpLSA-L model.

3.2. Document projection

In this step, the LR input test image I_{TST} is represented in the previously learnt LR topic space Θ_L. Initially, I_{TST} is interpolated to the target resolution as $\tilde{I}_{TST}$. Then, documents are extracted following the aforementioned image patch characterisation scheme. Finally, the sMpLSA-tst model (Fig. 2c) is used to estimate the $\Phi_{tra} \sim p(z|d)$ parameter considering Θ_L fixed. That is, the EM process (Algorithm 1) takes $n(w_L{,}d_{tst})$, K, λ and $p(w_L|z)$ as input arguments and the M-step only estimates the $p(z|d_{tst})$ parameter. Note that λ represents the sparsity factor of the Φ_{tst} distribution.

3.3. Topic-based SR

The target in this step is to reconstruct an initial super-resolved result I^*_{SR} following the sMpLSA model generative scheme. To achieve this goal, we initially provide a guess of the probability distribution $p(w_L|d_{tst})$ that each $\tilde{I}_{TST}$ test input patch d_{tst} belongs to the HR vocabulary w_H. This estimation can be easily worked out by marginalizing the sMpLSA model over topics as follows,

$$p(w_H|d_{tst}) = \frac{p(w_H{,}d_{tst})}{p(d_{tst})} = \sum_z p(w_H|z)p(z|d_{tst}) = \Theta_H \Phi_{tst}. \tag{17}$$

Note that this distribution provides probability values but word-counts are required to reconstruct the super-resolved gray levels values. Therefore, we use the number of words in each $\tilde{I}_{TST}$ patch, represented by the δ_{tst} prior term, to estimate the output number of words,

$$n(w_H|d_{tst}) = p(w_H|d_{tst})\delta_{tst} = p(w_H|d_{tst})\sum_{w_L} n(w_L|d_{tst}). \tag{18}$$

Finally, we reconstruct I^*_{SR} using a Gaussian-like windowing function [36] to alleviate possible misregistration effects when reconstructing the image from nearby overlapping patches. That is, a Gaussian kernel is initially applied to each document (image patch). Then, the corresponding document word-counts (gray level values) of the overlapping areas are averaged at each output pixel position. Eq. (19) represents this process where the $\mathcal{W}$ operator averages the document-word contributions to the final image pixel positions by means of a Gaussian kernel with $\sigma = 1$ standard deviation.

$$I^*_{SR} = \mathcal{W}(n(w_H|d_{tst})). \tag{19}$$

3.4. Post-processing

When considering a patch-based learning scheme, each patch is independently super-resolved and this may generate small pixel value discrepancies among patches in the final result, especially when the SR process is not conducted in the original image space. Precisely, this is the case of many manifold-based approaches and also the case of the proposed approach. In this situation, it is possible to use a final post-processing step [37] in order to guarantee a super-resolved output with the same pixel intensity value range of the LR input image.

As a result, the final stage is a post-processing step, based on the single-image Iterative Back Projection (IBP) approach [5], in order to mitigate possible deviations between the LR observation I_{TST} and the super-resolved result I^*_{SR}. Note that the proposed sMpLSA model provides an estimation of $p(w_H|d_{tst})$ (Eq. (17)), therefore the super-resolved patches are normalised as probability distributions. Precisely, this is the reason why we make use of the prior δ_{tst} to estimate the output pixel intensity values as Eq.(18) shows. However, this estimation may introduce some pixel value discrepancies among patches due to the fact that real HR word-counts are logically unknown. Eq. (20) illustrates the post-processing process.

$$I_{SR} = \underset{I_{SR}}{\operatorname{argmin}} \; \| \, \mathcal{D} \, \mathcal{B} \, I_{SR} - I_{TST} \, \|_2 \; + \; \alpha \, \| \, I_{SR} - I^*_{SR} \, \|_2 \tag{20}$$

$\mathcal{D}$ and $\mathcal{B}$ represent the decimating and blurring operators respectively. I^*_{SR} is the initial super-resolved result provided by Eq. (19) and I_{SR} is the final super-resolved output. Throughout this iterative process, the reconstruction error between the LR image I_{TST} and a simulated low-resolution version of the current estimate of the super-resolved image I_{SR} is minimised in order to obtain the final output result which guarantees a global reconstruction constraint. Note that Eq. (20) balances both the fitting of the final output image with the initial LR input and the fitting of the solution with itself by a factor α.

3.5. Computational complexity

Regarding the computational cost of the proposed TSR framework, we have to take into account two different complexities: the training cost (Sec. 3.1) and the test computational burden (Sec. 3.2-3.4). Since the latter is the actual cost required to super-resolve LR input images, we focus this analysis just on the test cost. In particular, three different operations are involved: the document projection (sMpLSA-tst), the topic-based SR (Eq.(19)) and the post-processing (Eq.(20)).

According to the standard pLSA model complexity [38], sMpLSA-tst cost is $\mathcal{O}(IN_HMK)$, where I is the maximum EM iterations, N_H represents the size of the HR vocabulary, M is the number of documents and K represents the number of topics. The computational burden of the topic-based SR process, conducted according to Eq.(19), is essentially $\mathcal{O}(N_HMK)$. Finally, the post-processing step has a total cost of $\mathcal{O}(I'N_HM)$ where I' represents the number of back-projection iterations that we fix at 100. As a result, the final computational cost of the TSR test phase is $\mathcal{O}(IN_HMK)$, that is, the computational burden of the regular pLSA model. However, it is important to highlight that the sMpLSA-tst model only has a single parameter to be estimated, i.e. Φ_{tst}, therefore it is expected to converge faster in practice.

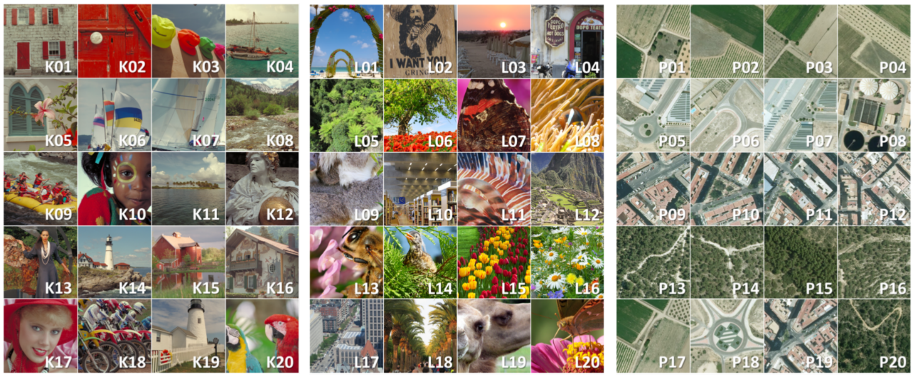
4. Experiments

The experiments presented here are aimed at validating the proposed approach performance against several LE-based SR algorithms available in the literature. In particular, Section 4.1 introduces the image datasets used in the experiments, Section 4.2 describes the experimental setting and Section 4.3 shows the obtained results.

4.1. Datasets

Figure 4 shows the three image datasets used in this work to conduct the experiments: (a) Kodak-20, a subset of 20 images from the Kodak Photo CD

PCD0992 collection [39], (b) the L-20 dataset presented in [37] and (c) PNOA-20, the remote sensing dataset proposed in [40]. We have considered a HR image size of 512 × 512 pixels, therefore datasets have been pre-processed accordingly. Specifically, Kodak-20 images have been cropped to 512 × 512 pixels, L-20 images have been down-scaled to the considered HR size via the Matlab R2016b *imresize*[1] function and images from PNOA-20 dataset do not require any kind of pre-processing.



(a) Kodak-20 (b) L-20 (c) PNOA-20

Figure 4: Image databases used in the experiments. The first sixteen images (form 01 to 16) are used for training purpose and the last four (from 17 to 20) serve as the test set.

Once the datasets' HR images have been created, the Matlab R2016b *imresize* function has been also used to generate the corresponding LR images according to the considered scaling factors.

4.2. Experimental settings

The proposed approach has been validated against 7 different reference LE-based SR methods selected from the literature. In particular, we have chosen for comparison purposes one sparse coding method, VSR [10], two neighbourhood embedding approaches, ANR+ [41] and GLR [11], and four mapping methods, namely CNN [42], JOR [17], SRF [43] and LKR [44]. Additionally, we use the bi-cubic interpolation kernel (BCI) as the baseline assessment method.

[1] By default, this function performs anti-aliasing when shrinking an image by applying a scaled version of the bi-cubic interpolation kernel.

All these reference methods have been selected because their implementations are publicly available and besides they tend to introduce some kind of image semantics along the SR process [16]. With the exception of VSR and CNN, which represent the most classical sparse coding and deep learning-based approaches, each one of the tested methods uses a particular scheme to take advantage of the image semantics when super-resolving images. ANR+ and GLR use a correlation-based clustering process over trained dictionary atoms to learn multiple patch embeddings. JOR performs an EM clustering over training patches to learn a different mapping function for each cluster. SFR introduces an ℓ^2-based regularisation term when learning the tree structure in order to grantee similar patches on leaves. LKR uses the k-means algorithm over dictionary atoms to train several kernel regressors.

Experiments have been conducted considering two different scaling factors, 2× and 4×, in order to achieve a super-resolve output with a size of 512×512 pixels. For each one of the three considered datasets (Fig. 4), the first sixteen images (from 01 to 16) have been used as a training set and the last four images (from 17 to 20) have been employed as test. Note that three-quarters of the data are considered for training, which is a common scenario for hold-out validation in machine learning algorithms. Besides, the use of this configuration over the three considered datasets also guarantees a high data diversity when validating the considered learning-based models. In particular, all the SR methods have been trained for each dataset using a subset of 100,000 patches and their corresponding default settings for their algorithm parameters.

Regarding the proposed approach (TSR), we have followed a similar settings to the ones presented in [34]. In particular, a patch size $s = 15$, a number of topics $K = 1000$, a post-processing step with a Gaussian blurring operator ($\sigma = 0.6$) together with 100 back-projection iterations and a sparsity factor $\lambda = 1$. Note that we use the λ factor to control the entropy of the $p(z|d)$ and $p(z|d)_{tst}$ probability distributions. Specifically, the second term of Eq. 11 and Eq. 15) deactivates the topic-document components (i.e. image patterns associated to a given image patch) with a probability value lower than λ/K. As a result, $\lambda = 1$ allows neglecting the components under the probability of the uniform distribution $1/K$ which is the most uninformative configuration. In order to perform the comparison as fair as possible, the number of atoms in sparse coding and neighbourhood embedding methods have been fixed to $K = 1000$ due to the fact that the number of topics plays a similar role. That is, the K parameter in TSR represents the amount of hidden patterns used to represent the data. Therefore, this value is comparable to the number of dictionary atoms considered in a sparse coding-based approach or to the number of neighbours used in a neighbourhood embedding method mainly because they all define the number of different components considered when super-resolving patches.

In this work, two reference metrics are used to assess the quality of the super-resolved images, PSNR (Peak Signal to Noise Ratio) [2] and SSIM (Structural SIMilarity) [45]. On the one hand, PSNR measures the difference between the maximum power of the ground-truth image and the noise appearing in the super-resolved result. On the other hand, SSIM evaluates the correlation, intensity and contrast of the super-resolved image with respect to its ground-truth counterpart. Note that the higher the PSNR and SSIM values, the better the quality of the super-resolved result. Finally, it should be mentioned that a 7-pixel security image border has been discarded when computing these metrics, due to the fact that patch overlapping in the image borders is imprecise because partial neighbour information is not available.

4.3. Results

Tables 1-2 present the assessment of the super-resolved test images for Kodak-20, L-20 and PNOA-20 datasets in terms of the PSNR and SSIM metrics. Specifically, Table 1 contains the results when considering a 2× scaling factor and Table 2 the corresponding results for a 4× factor.

The super-resolution methods used in this work are shown in columns, that is, first the BCI baseline interpolation, subsequently the seven LE-based SR methods extracted from the literature (VSR, ANR+, GLR, CNN, JOR, SRF and LKR) and finally the proposed approach (TSR). In rows, we show for each test image of each database its corresponding SR assessment in terms of the PSNR and SSIM metrics. Note that the last row provides the methods' average computational time.

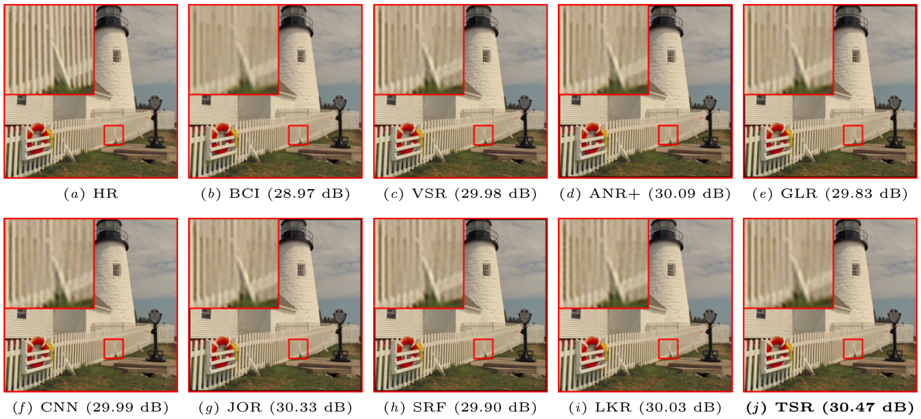
In addition to the quantitative evaluation provided by the PSNR and SSIM metrics, some visual results are provided as a qualitative evaluation for the tested SR methods. Specifically, Figures 5-6 show the super-resolved results obtained for K19 and P20 test images considering a 2× scaling factor. Besides, Figure 7 and Figure 8 present the results for K20 and L17 test images with a 4× scaling factor

Table 1: SR quality assessment for Kodak-20, L-20 and PNOA-20 datasets considering a 2× scaling factor. In rows, super-resolved test images and metrics, PSNR (db) and SSIM. In columns, the tested SR methods including the prosed approach (last column). The best result for each row is highlighted in bold. Note that the last row shows the methods' average computational time.

Database	Training set	Test Image	Quality Metric	SR Methods								
				BCI	VSR	ANR+	GLR	CNN	JOR	SRF	LKR	TSR
Kodak-20	K01-K16	K17	SSIM	*0.969*	*0.976*	*0.976*	*0.975*	*0.975*	*0.976*	*0.975*	*0.976*	***0.976***
			PSNR (dB)	*34.74*	*35.68*	***35.84***	*35.59*	*35.45*	*35.76*	*35.55*	*35.70*	*35.55*
		K18	SSIM	*0.951*	*0.971*	*0.971*	*0.967*	*0.966*	*0.971*	*0.968*	*0.969*	***0.972***
			PSNR (dB)	*27.11*	*28.68*	***29.00***	*28.33*	*28.19*	*28.83*	*28.49*	*28.63*	*28.56*
		K19	SSIM	*0.936*	*0.953*	*0.953*	*0.951*	*0.956*	*0.954*	*0.951*	*0.952*	***0.965***
			PSNR (dB)	*28.97*	*29.98*	*30.09*	*29.83*	*29.99*	*30.33*	*29.90*	*30.03*	***30.47***
		K20	SSIM	*0.982*	***0.987***	*0.987*	*0.986*	*0.985*	*0.987*	*0.986*	*0.986*	*0.986*
			PSNR (dB)	*35.12*	*36.42*	***36.61***	*36.17*	*35.98*	*36.60*	*36.19*	*36.50*	*36.13*
L-20	L01-L16	L17	SSIM	*0.926*	*0.952*	*0.953*	*0.948*	*0.955*	*0.953*	*0.949*	*0.950*	***0.965***
			PSNR (dB)	*24.94*	*25.83*	*26.03*	*25.76*	*26.07*	*26.07*	*26.02*	*25.91*	***26.53***
		L18	SSIM	*0.926*	*0.946*	*0.947*	*0.942*	*0.947*	*0.947*	*0.944*	*0.945*	***0.955***
			PSNR (dB)	*27.12*	*27.86*	*27.96*	*27.70*	*27.87*	*27.97*	*27.79*	*27.85*	***28.17***
		L19	SSIM	*0.934*	*0.946*	*0.946*	*0.944*	*0.950*	*0.946*	*0.944*	*0.945*	***0.956***
			PSNR (dB)	*29.70*	*30.18*	*30.23*	*30.11*	*30.26*	*30.23*	*30.13*	*30.15*	***30.48***
		L20	SSIM	*0.976*	*0.983*	*0.983*	*0.982*	*0.982*	*0.983*	*0.981*	*0.982*	***0.984***
			PSNR (dB)	*33.20*	*34.45*	***34.65***	*34.17*	*34.24*	*34.63*	*34.15*	*34.39*	*34.55*
PNOA-20	P01-P16	P17	SSIM	*0.897*	*0.904*	*0.911*	*0.904*	*0.910*	*0.906*	*0.899*	*0.897*	***0.926***
			PSNR (dB)	*29.37*	*30.04*	*30.20*	*29.85*	*29.94*	*30.06*	*29.84*	*29.72*	***30.27***
		P18	SSIM	*0.972*	*0.983*	*0.983*	*0.981*	*0.982*	*0.983*	*0.981*	*0.982*	***0.984***
			PSNR (dB)	*33.54*	*34.56*	***34.71***	*34.48*	*34.34*	*34.70*	*34.49*	*34.64*	*34.58*
		P19	SSIM	*0.976*	***0.987***	*0.987*	*0.985*	*0.984*	*0.987*	*0.985*	*0.987*	*0.985*
			PSNR (dB)	*31.42*	*32.92*	*33.16*	*32.79*	*32.30*	*33.17*	*32.87*	***33.21***	*32.40*
		P20	SSIM	*0.950*	*0.968*	*0.967*	*0.965*	*0.970*	*0.967*	*0.966*	*0.967*	***0.975***
			PSNR (dB)	*28.94*	*29.71*	*29.78*	*29.66*	*29.95*	*29.76*	*29.73*	*29.82*	***30.32***
Average time (s)				***0.05***	*493.71*	*1.82*	*30.86*	*5.55*	*869.51*	*42.99*	*84.25*	*388.31*

Table 2: SR quality assessment for Kodak-20, L-20 and PNOA-20 datasets considering a 4× scaling factor. In rows, super-resolved test images and metrics, PSNR (db) and SSIM. In columns, the tested SR methods including the prosed approach (last column). The best result for each row is highlighted in bold. Note that the last row shows the methods' average computational time.

Database	Training set	Test Image	Quality Metric	SR Methods								
				BCI	VSR	ANR+	GLR	CNN	JOR	SRF	LKR	TSR
Kodak-20	K01-K16	K17	SSIM	*0.905*	*0.912*	*0.916*	*0.915*	*0.912*	*0.916*	*0.912*	*0.913*	***0.919***
			PSNR (dB)	*30.39*	*30.86*	*31.12*	*30.99*	*30.71*	***31.18***	*30.93*	*30.93*	*30.95*
		K18	SSIM	*0.813*	*0.848*	*0.857*	*0.847*	*0.832*	***0.857***	*0.849*	*0.850*	*0.851*
			PSNR (dB)	*22.39*	*22.88*	*23.15*	*22.89*	*22.73*	***23.18***	*22.90*	*22.97*	*23.10*
		K19	SSIM	*0.814*	*0.831*	*0.837*	*0.832*	*0.829*	*0.839*	*0.832*	*0.831*	***0.846***
			PSNR (dB)	*24.49*	*24.82*	*24.97*	*24.84*	*24.86*	*25.06*	*24.83*	*24.83*	***25.25***
		K20	SSIM	*0.934*	*0.940*	*0.944*	*0.942*	*0.940*	*0.945*	*0.940*	*0.940*	***0.947***
			PSNR (dB)	*29.44*	*29.95*	*30.23*	*30.07*	*29.85*	***30.28***	*29.97*	*29.98*	*30.26*
L-20	L01-L16	L17	SSIM	*0.756*	*0.791*	*0.799*	*0.789*	*0.778*	*0.799*	*0.793*	*0.791*	***0.801***
			PSNR (dB)	*21.16*	*21.42*	*21.56*	*21.46*	*21.44*	*21.60*	*21.49*	*21.43*	***21.71***
		L18	SSIM	*0.797*	*0.826*	*0.831*	*0.824*	*0.814*	***0.832***	*0.827*	*0.826*	*0.832*
			PSNR (dB)	*23.79*	*24.05*	*24.16*	*24.07*	*24.05*	*24.21*	*24.10*	*24.08*	***24.32***
		L19	SSIM	*0.849*	*0.864*	*0.866*	*0.863*	*0.859*	*0.866*	*0.863*	*0.863*	***0.868***
			PSNR (dB)	*27.43*	*27.60*	*27.70*	*27.67*	*27.62*	*27.73*	*27.59*	*27.62*	***27.77***
		L20	SSIM	*0.916*	*0.923*	*0.929*	*0.925*	*0.923*	*0.930*	*0.920*	*0.922*	***0.931***
			PSNR (dB)	*28.22*	*28.74*	*29.10*	*28.83*	*28.64*	***29.19***	*28.70*	*28.76*	*28.99*
PNOA-20	P01-P16	P17	SSIM	*0.800*	*0.824*	***0.827***	*0.821*	*0.812*	*0.827*	*0.824*	*0.823*	*0.827*
			PSNR (dB)	*26.50*	*26.85*	*26.98*	*26.86*	*26.72*	*26.96*	*26.87*	*26.90*	***26.98***
		P18	SSIM	*0.875*	*0.900*	*0.901*	*0.897*	*0.889*	***0.905***	*0.898*	*0.898*	*0.903*
			PSNR (dB)	*28.29*	*28.93*	*28.96*	*28.92*	*28.71*	***29.16***	*28.91*	*28.92*	*29.14*
		P19	SSIM	*0.864*	*0.897*	*0.909*	*0.894*	*0.880*	***0.912***	*0.895*	*0.896*	*0.896*
			PSNR (dB)	*24.67*	*25.58*	*26.06*	*25.60*	*25.16*	***26.20***	*25.68*	*25.70*	*25.65*
		P20	SSIM	*0.802*	*0.840*	*0.841*	*0.836*	*0.823*	*0.841*	*0.841*	*0.840*	***0.842***
			PSNR (dB)	*24.43*	*24.68*	*24.71*	*24.73*	*24.77*	*24.77*	*24.74*	*24.73*	***25.12***
Average time (s)				***0.04***	*469.90*	*0.74*	*8.22*	*5.56*	*317.55*	*14.04*	*91.00*	*355.97*


(*a*) HR (*b*) BCI (28.97 dB) (*c*) VSR (29.98 dB) (*d*) ANR+ (30.09 dB) (*e*) GLR (29.83 dB)

(*f*) CNN (29.99 dB) (*g*) JOR (30.33 dB) (*h*) SRF (29.90 dB) (*i*) LKR (30.03 dB) **(*j*) TSR (30.47 dB)**

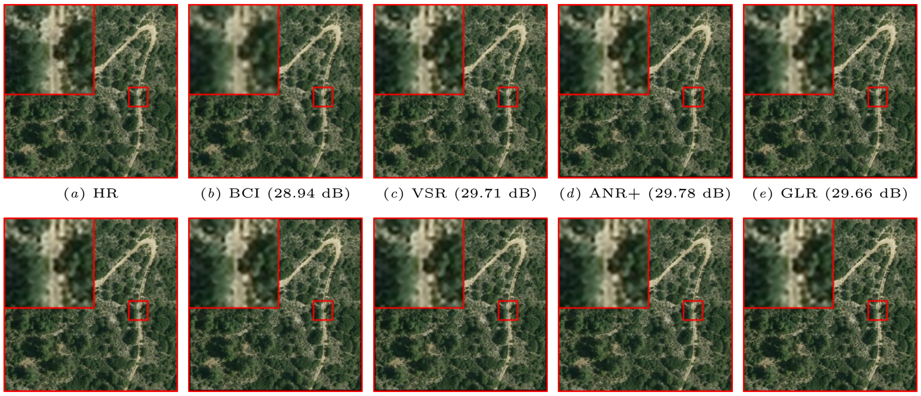
Figure 5: SR results obtained using the methods shown in captions over the test image K19 with a 2× scaling factor. For each result, PSNR (dB) values appear in brackets. The best PSNR value is highlighted in bold.

5. Discussion

The quantitative assessment reported in Tables 1-2 show how the proposed approach is able to achieve a competitive performance in the three considered datasets. When considering a 2× scaling factor (Table 1), the proposed approach TSR together with the mapping method JOR and the neighbourhood embedding technique ANR+ obtain, on average, the three best SSIM values. In the case of the PSNR metric, four methods deserve to be mentioned: ANR+, TSR and JOR. The first one (ANR+) obtains the best PSNR value for Kodak-20 and PNOA-20 collections whereas the proposed approach (TSR) does the same for L-20. In the case of JOR, this method deserve to be in the third overall place.

A similar trend can be observed when considering a 4× scaling factor (Table 2). In this case, JOR obtains the best average SSIM value in PNOA-20 while TSR achieves the best result in Kodak-20 and L-20. Besides, ANR+ obtains the third best average SSIM value. Regarding the PSNR metric, JOR obtains the best average result in PNOA-20 and TSR reaches the best performance in Kodak-20 and L-20. Finally, ANR+ achieves the third best PSNR result on average.

Overall, JOR and TSR methods have shown to obtain the best quantitative performance followed some way behind by ANR+. However, differences between JOR and TSR are relatively small what motivates a thorough discussion over



(*a*) HR (*b*) BCI (28.94 dB) (*c*) VSR (29.71 dB) (*d*) ANR+ (29.78 dB) (*e*) GLR (29.66 dB)

(*f*) CNN (29.95 dB) (*g*) JOR (29.76 dB) (*h*) SRF (29.73 dB) (*i*) LKR (29.82 dB) **(*j*) TSR (30.32 dB)**

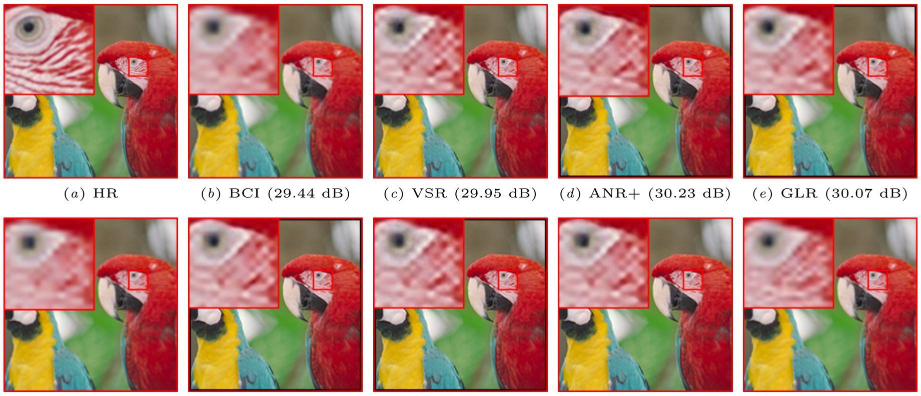
Figure 6: SR results obtained using the methods shown in captions over the test image P20 with a 2× scaling factor. For each result, PSNR (dB) values appear in brackets. The best PSNR value is highlighted in bold.

qualitative results to find out methods' singularities.

According to the visual results presented in Figures 5-8, each SR method tends to foster a particular kind of visual features on the super-resolved output. Some methods, like JOR or LKR, are able to obtain sharper edges, while others, like VSR or ANR+, seem more robust to noise by generating smoother super-resolved textures.

In terms of visual perceived quality, the proposed approach (TSR) achieves a remarkable performance. For instance, the fence detail in Fig. 5(j) is certainly the most similar to its HR counterpart in Fig. 5(a). Even though the result provided by JOR (Fig. 5(g)) seems to obtain a slightly better contrast on some parts of the image, the proposed approach is able to introduce more high-frequency information in the fence structure. Another illustrative example can be found in Fig. 6 where it is possible to see that the proposed approach introduces some fine details in the vegetation which are not present in other methods' results.

When considering a 4× scaling factor, the proposed approach also shows its capability to recover high-frequency details, however some other SR methods seem to generate more image contrast. For instance, it is the case of the result provided by JOR in Fig. 7.(g) which achieves a great visual performance providing sharp edges. Nonetheless, it generates a kind of watering effect and also increases the aliasing on the output image. In the proposed approach result



(*a*) HR (*b*) BCI (29.44 dB) (*c*) VSR (29.95 dB) (*d*) ANR+ (30.23 dB) (*e*) GLR (30.07 dB)

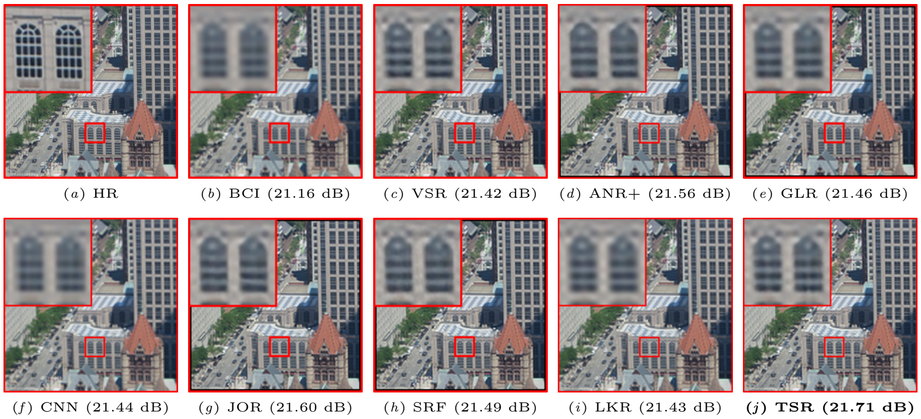
(*f*) CNN (29.85 dB) **(*g*) JOR (30.28 dB)** (*h*) SRF (29.97 dB) (*i*) LKR (29.98 dB) (*j*) TSR (30.26 dB)

Figure 7: SR results obtained using the methods shown in captions over the test image K20 with a 4× scaling factor. For each result, PSNR (dB) values appear in brackets. The best PSNR value is highlighted in bold.

shown in Fig. 7.(j), we can see that edges are not so contrasted but the aliasing distortion is slightly reduced while new high-frequency information of the stripe pattern is recovered. A similar behaviour can be observed in the window detail of Fig. 8. In this case, the proposed approach (Fig. 7.(j)) seems to recover the vertical pattern of the window better than JOR (Fig. 7.(g)).

Regarding the computational time, we can observe important differences among the tested methods. In particular, four algorithm groups can be identified when super-resolving LR input images: (i) BCI and ANR+, with an average time consumption per image under 3 seconds, (ii) GLR CNN and SRF, with a time between 10 and 60 seconds, (iii) LKR, which require between 60 and 120 seconds and (iv) VSR, JOR and TSR, with a computational time between 300 and 500 seconds. The proposed approach is definitely not one of the most computationally efficient methods, however it is able to obtain a computational cost similar to that of JOR which has shown to be one of the best methods.

Finally, another noteworthy point is related to the use of the post-processing step. As it was mentioned in Sec. 3.4, the proposed approach is able to take advantage of the IBP process in order to relieve some possible pixel value deviations generated in the $n(w_H|d_{tst})$ estimation. In particular, the PSNR gain obtained by TSR when using the specified post-processing step is, on average, 0.05 dB. Additionally, we have also tested that the JOR method does not obtain, on average, a performance improvement when using such process.



(*a*) HR (*b*) BCI (21.16 dB) (*c*) VSR (21.42 dB) (*d*) ANR+ (21.56 dB) (*e*) GLR (21.46 dB)

(*f*) CNN (21.44 dB) (*g*) JOR (21.60 dB) (*h*) SRF (21.49 dB) (*i*) LKR (21.43 dB) **(*j*) TSR (21.71 dB)**

Figure 8: SR results obtained using the methods shown in captions over the test image L17 with a 4× scaling factor. For each result, PSNR (dB) values appear in brackets. The best PSNR value is highlighted in bold.

5.1. Proposed approach advantages and limitations

When comparing the TSR results to the ones obtained by the other semantic-based SR methods, we can observe the proposed approach potential. Even though the straightforward clustering approach is the most extended way to introduce image semantics in the SR process, the effectiveness of this approach is limited by the intra-cluster semantic variability. Note that a clustering process naturally tends to group similar patches within the same cluster. However, the inherent information loss in the LR domain may produce that two patches related to two completely different semantic concepts become part of the same cluster.

In order to overcome the above mentioned limitation, the proposed approach works by super-resolving latent patterns instead of image patches themselves. That is, the SR process is driven by the mixture of latent patterns appearing in the LR input image and this allows TSR to recover a richer variety of high-frequency patterns for a given LR patch. In a sense, the proposed method provides a more flexible scheme than the current semantic-based SR techniques because LR patches are allowed to have simultaneously multiple SR paths through the latent patterns defined by topics and therefore more HR patterns can be involved in the SR process.

However, this higher flexibility has a main implication: a blurring effect may appear if too many HR patterns are involved. In order to reduce this possible effect, we introduce the λ sparsity constraint to control the number of considered HR patterns when super-resolving LR images. In spite of this, it may be difficult to find the ideal sparsity factor because it logically depends on the input image features as well as the considered scaling factor. In this work, we assume a constant λ factor to define the sMpLSA model but further research could be directed to this extent.

6. Conclusions and future work

In this work, we presented a topic-based SR framework in order to super-resolve LR images according to the semantic patterns encapsulated by the latent topic space. Specifically, we initially define the sMpLSA model and then we used this model to super-resolve LR images by super-resolving latent topics instead of image patches themselves. Finally, we conducted an experimental comparison over three different image datasets to show the proposed approach performance with respect to different reference LE-based SR methods available in the literature.

One of the main conclusions that arises from this work is the potential of topic models to cope with the SR problem because of their capabilities to manage data semantics. Whereas the common SR trend relies on using a clustering-based process in the image patch representation to define the image semantics, we proposed to transform this classical perspective into a new probabilistic approach where the SR process can be performed using the semantics encapsulated by the sMpLSA model in the latent topic space.

According to the conducted experiments, the proposed approach obtains a competitive performance over the three considered databases in terms of both quantitative and qualitative results. Regarding the SSIM and PSNR metrics, the SR framework proposed in this work obtains, on average, a similar performance to the one obtained by the mapping approach JOR. Besides, it is able to outperform the rest of the tested methods. Considering the visual results, the proposed approach has shown to be one of the most effective methods especially when considering a 2× scaling factor.

Although the proposed approach results are encouraging as a semantic-based SR technique, it still has some limitations which provide room for improvement to conduct more research on topic-based SR. Specifically, future work is aimed at the following directions: (i) a sMpLSA extension to estimate the ideal sparsity factor for each input patch, (ii) automatic procedures to set the most appropriate number of topics and (iii) 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 projects ESP2013-48458-C4-3-P and ESP2016-79503-C2-2-P, by Generalitat Valenciana through the PROMETEO-II/2014/062 project and the APOSTD/2017/007 contract, and by Universitat Jaume I under the P11B2014-09 project.

References

[1] J. van Ouwerkerk, Image super-resolution survey, Image and Vision Computing 24 (10) (2006) 1039–1052.

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

[3] C.-Y. Yang, C. Ma, M.-H. Yang, Single-image super-resolution: A benchmark, in: European Conference on Computer Vision, 2014.

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

[5] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graph. Models Image Process. 53 (3) (1991) 231–239.

[6] J. Sun, Z. Xu, H.-Y. Shum, Image super-resolution using gradient profile prior, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[7] Q. Shan, Z. Li, J. Jia, C.-K. Tang, Fast image/video upsampling, ACM Transactions on Graphics (SIGGRAPH ASIA).

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

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

[10] J. Yang, J. Wright, T. S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Img. Proc. 19 (11) (2010) 2861–2873.

[11] R. Timofte, V. De Smet, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in: IEEE International Conference on Computer Vision, 2013, pp. 1920–1927.

[12] C. Dong, C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, Vol. 8692, 2014, pp. 184–199.

[13] G. Polatkan, M. Zhou, L. Carin, D. Blei, I. Daubechies, A bayesian nonparametric approach to image super-resolution, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (2) (2015) 346–358.

[14] D. Glasner, S. Bagon, M. Irani, Super-resolution from a single image, in: IEEE International Conference on Computer Vision, 2009.

[15] T. Michaeli, M. Irani, Blind deblurring using internal patch recurrence, in: European Conference on Computer Vision, 2014, pp. 783–798.

[16] R. Timofte, V. D. Smet, L. V. Gool, Semantic super-resolution: When and where is it useful?, Computer Vision and Image Understanding 142 (2016) 1–12.

[17] D. Dai, R. Timofte, L. Van Gool, Jointly optimized regressors for image super-resolution, in: Eurographics, 2015.

[18] D. M. Blei, Probabilistic topic models, Communications of the ACM 55 (4) (2012) 77–84.

[19] D.-T. Vo, C.-Y. Ock, Learning to classify short text from scientific documents using topic models with various types of knowledge, Expert Systems with Applications 42 (3) (2015) 1684–1698.

[20] R. Fernandez-Beltran, R. Montoliu, F. Pla, Vocabulary reduction in bow representing by topic modeling, in: Iberian Conference on Pattern Recognition and Image Analysis, 2013, pp. 648–655.

[21] R. Fernandez-Beltran, F. Pla, Latent topic encoding for content-based retrieval, in: Iberian Conference on Pattern Recognition and Image Analysis, 2015, pp. 640–648.

[22] S. Nikolopoulos, S. Zafeiriou, I. Patras, I. Kompatsiaris, High order plsa for indexing tagged images, Signal Processing 93 (8) (2013) 2212–2228.

[23] R. Fernandez-Beltran, F. Pla, Latent topics-based relevance feedback for video retrieval, Pattern Recognition 51 (2016) 72–84.

[24] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, Machine Learning 42 (1-2) (2001) 177–196.

[25] D. Blei, A. Ng, M. Jordan, Latent dirichlet allocation, Journal of Machine Learning Research 3 (4-5) (2003) 993–1022.

[26] J. Varadarajan, J. M. Odobez, Topic models for scene analysis and abnormality detection, in: IEEE International Conference on Computer Vision Workshops, 2009, pp. 1338–1345.

[27] Z. Lyu, S. Zhao, S. Fang, H. Liang, Multi image super resolution reconstruction using a novel degradation model, in: International Conference on Audio, Language and Image Processing, 2014.

[28] O. Isupova, D. Kuzin, L. Mihaylova, Learning methods for dynamic topic modeling in automated behavior analysis, IEEE Transactions on Neural Networks and Learning Systems PP (99) (2017) 1–14.

[29] D. Pathak, A. Sharang, A. Mukerjee, Anomaly localization in topic-based analysis of surveillance videos, in: 2015 IEEE Winter Conference on Applications of Computer Vision, 2015, pp. 389–395.

[30] S. Romberg, R. Lienhart, E. Hörster, Multimodal image retrieval, International Journal of Multimedia Information Retrieval 1 (1) (2012) 31–44.

[31] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-graber, D. M. Blei, Reading tea leaves: How humans interpret topic models, in: Advances in Neural Information Processing Systems, 2009, pp. 288–296.

[32] X. Chen, A. Choudhury, P. van Beek, A. Segall, Facial video super resolution using semantic exemplar components, in: IEEE International Conference on Image Processing, 2015, pp. 1314–1318.

[33] P. Purkait, B. Chanda, Image upscaling using multiple dictionaries of natural image patches, in: Asian Conference on Computer Vision, Vol. 7726, 2013, pp. 284–295.

[34] R. Fernandez-Beltran, P. Latorre-Carmona, F. Pla, Latent topic-based super-resolution for remote sensing, Remote Sensing Letters 8 (6) (2017) 498–507.

[35] Y. Zhang, R. Jin, Z. Zhou, Understanding bag-of-words model: a statistical framework, Int. J. Mach. Learn. Cyber. 1 (1) (2010) 43–52.

[36] S. Alliney, C. Morandi, Digital image registration using projections, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (2) (1986) 222–233.

[37] R. Timofte, R. Rothe, L. V. Gool, Seven ways to improve example-based single image super resolution, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[38] R. Fernandez-Beltran, F. Pla, Incremental probabilistic latent semantic analysis for video retrieval, Image and Vision Computing 38 (2015) 1–12.

[39] Kodak photo cd pcd0992, `http://r0k.us/graphics/kodak/`, accessed: 2016-04-06 (2016).

[40] R. Fernandez-Beltran, P. Latorre-Carmona, F. Pla, Single-frame super-resolution in remote sensing: a practical overview, International Journal of Remote Sensing 38 (1) (2017) 314–354.

[41] R. Timofte, V. D. Smet, L. V. Gool, A+: Adjusted anchored neighborhood regression for fast super-resolution, in: ACCV, 2014.

[42] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2) (2016) 295–307.

[43] S. Schulter, C. Leistner, H. Bischof, Fast and accurate image upscaling with super-resolution forests, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[44] R. Liao, Z. Qin, Image super-resolution using local learnable kernel regression, in: Asian Conference on Computer Vision, 2012.

[45] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.