Signal Processing 152 (2018) 227–237

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Sparse multi-modal probabilistic latent semantic analysis for single-image super-resolution

Ruben Fernandez-Beltran*, Filiberto Pla

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

ARTICLE INFO

Article history:
Received 1 August 2017
Revised 23 January 2018
Accepted 27 May 2018

Keywords:
Super-Resolution
Latent topics
Probabilistic latent semantic analysis
Image learning
Image quality assessment

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.

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 (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 [1–4]. 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

* Corresponding author.
E-mail addresses: rufernan@uji.es (R. Fernandez-Beltran), pla@uji.es (F. Pla).

https://doi.org/10.1016/j.sigpro.2018.05.026
0165-1684/

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 $D=\{d_1,d_2,\ldots,d_M\}$ in a particular word-space $W=\{w_1,w_2,\ldots,w_N\}$, 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 [26–30], 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 $Z=\{z_1,z_2,\ldots,z_K\}$. 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 [16,32], 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 en-