INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)
Available online at www.ijariit.com

A Survey on Sparsity Based Single Image Super Resolution of Colour Images

Helena Thomas
helenathomas2007@gmail.com
College of Engineering Kidangoor, Kottayam, Kerala

Anitha .R
anitharshibu@gmail.com
College of Engineering Kidangoor, Kottayam, Kerala

ABSTRACT

There are several methods for improving the resolution of images. The usual approach is by sparsely representing patches in a low-resolution input image via a dictionary of an example low resolution patches and then using the coefficients of this representation to generate the high-resolution output via an analogous high-resolution dictionary. However, most existing methods focus on luminance channel information only and neglect the colour channels. The present method achieves sparsity based super resolution by considering multiple colour channels also along with luminance channel Information. Edge similarities among RGB colour bands are used as cross channel correlation constraints. A dictionary learning method specifically to learn colour dictionaries that encourage edge similarities is also used. The advantages of this method are demonstrated both visually and quantitatively using image quality metrics.

Keywords: *Colour Super Resolution, Single-image super Resolution, Sparse Coding, Dictionary Learning, Edge Similarity.*

1. INTRODUCTION

Super Resolution (SR) of an image is the process of obtaining a high-resolution image by enhancing a low-resolution input image. Conventional Super-Resolution (SR) approaches maps multiple Low Resolution (LR) images of the same scene as input to a High Resolution (HR) image based on some reasonable assumptions, prior knowledge, or capturing the diversity in LR images.

The Super Resolution task encounters an issue that much information is lost in the process of going from high resolution images to low resolution images and hence the solution is not unique. Consequently, strong prior information is incorporated to yield realistic and robust solutions. Example priors include knowledge of the underlying scene, distribution of pixels, historical data, smoothness and edge information and so on.

In contrast to conventional super resolution problem with multiple low-resolution images as input, single image super resolution methods have been developed recently that generate the high-resolution image only based on a single low-resolution image. Classically, the solution to this problem is based on example-based methods exploiting nearest neighbour estimations, where pairs of low and high-resolution image patches are collected, and each low-resolution patch is mapped to a corresponding high-resolution patch.

Freeman et al. proposed an estimation scheme where high-frequency details are obtained by taking nearest neighbour based estimation on low resolution patches. Glasner et al. used the observation that patches in a natural image tend to redundantly recur many times inside the image, both within the same scale, as well as across different scales and approached the single image super resolution problem.

An alternate mapping scheme was proposed by Kim et al. using kernel ridge regression. Many learning techniques have been developed which attempt to capture the co-occurrence of low resolution and high-resolution image patches. Inspired by manifold forming methods like locally linear embedding (LLE), Chang et al. proposed a neighbourhood embedding approach.

Specifically, small image patches in the low and high-resolution images form manifolds with similar local geometry in two distinct feature spaces and local geometry information is used to reconstruct a patch using its neighbours in the feature space.

More recently, sparse representation based methods have been applied to the single image super resolution problem. Essentially in these techniques, a historical record of typical geometrical structures observed in images is exploited and examples of high and low-resolution image patches are collected as a dictionary (matrix). Yang et al. proposed to apply sparse coding for retrieving the high-resolution image from the LR image. Zeyde et al. extended this method to develop a local Sparse-Land model on image patches. Timofte et al. proposed the Anchored Neighbourhood Regression (ANR) method which uses learned dictionaries in combination with neighbour embedding methods.

Other super resolution methods based on statistical signal processing or dictionary learning methods have been proposed. On top of sparsity based methods, learning based methods have also been exploited for SR problem to learn dictionaries that are more suitable for this task. Mostly, dictionary learning or example-based learning methods in super-resolution use an image patch or feature-based approach to learn the relationship between high resolution scenes and their low-resolution counterparts.

Yang et al. propose to use a collection of raw image patches as dictionary elements in their framework. Subsequently, a method that learns LR and HR dictionaries jointly was proposed. A semi-coupled dictionary learning (SCDL) model and a mapping function were proposed where the learned dictionary pairs can characterize the structural features of the two image domains, while the mapping function reveals the intrinsic relationship between the two.

In addition, coupled dictionary learning for the same problem was proposed, where the learning process is modelled as a bilevel optimization problem. Dual or joint filter learning in addition to dual (joint) dictionaries was developed

2. LITERATURE REVIEW

In the following section we will be discussing various methods promoting safer transportation:

A. *Screen Content Image Segmentation Using Sparse Decomposition and Total Variation Minimization*:

This paper by the author Shervin Minaee proposed that sparse decomposition has been widely used for different applications, such as source separation, image classification, image denoising and more. The paper presents a new algorithm for segmentation of an image into background and foreground text and graphics using sparse decomposition and total variation minimization. The proposed method is designed based on the assumption that the background part of the image is smoothly varying and can be represented by a linear combination of a few smoothly varying basis functions, while the foreground text and graphics can be modelled with a sparse component over laid on the smooth background.

The background and foreground are separated using a sparse decomposition framework regularized with a few suitable regularization terms which promote the sparsity and connectivity of foreground pixels. This algorithm has been tested on a dataset of images extracted from HEVC standard test sequences for screen content coding, and is shown to have superior performance over some prior methods, including least absolute deviation fitting, k-means clustering based segmentation in DjVu filter and shape primitive extraction and coding (SPEC) algorithm.

The issues with previous algorithms motivate us to design a new segmentation algorithm which overcomes the problems of previous algorithms. We propose a sparse-decomposition framework to perform this image segmentation task. Sparse representation has been used for various applications in recent years, including face recognition, super-resolution, morphological component analysis, image classification, image restoration and sparse coding. Instead of looking at the intensities of pixels and deciding whether it should belong to background or foreground, we believe it is better to look at the smoothness of a group of pixels and then decide whether they should belong to background or foreground. The other important observation is that the foreground layer should contain a set of connected pixels (such as the pixels in a text stroke, or a line in graphics), not a set of randomly located isolated points. Therefore, in our segmentation algorithm, we enforce the extracted foreground pixels to be connected to each other by penalizing their total variation. Based on these two notions, we propose a sparse decomposition framework for the segmentation task. We model the background part of the image with a linear combination of a set of smoothly varying basis functions, and the foreground layer with a sparse component that has connected pixels.

Smooth background regions can be well represented with a few smooth basis functions, whereas the high-frequency component of the image belonging to the foreground, cannot be modelled with a smooth model. But using the fact that foreground pixels occupy a relatively small percentage of the images that can be modelled with a sparse component overlaid on the background. Therefore, it is natural to think of the mixed content image as a superposition of two components, one smooth and the other one sparse. Therefore, we can use signal decomposition techniques to separate these two components.

We first need to derive a suitable model for background component. We divide each image into non-overlapping blocks of size $N \times N$, and then represent each image block denoted by $F(x, y)$, with a smooth model $B(x, y; \alpha_1 \ldots, \alpha_K)$, where x and y denote the horizontal and vertical axes and $\alpha_1 \ldots, \alpha_K$ denote the parameters of this smooth model. For the choice of smooth model, we propose to use a linear combination of K basis functions $k=1\ \alpha_k P_k(x, y)$, where $P_k(x, y)$ denotes a 2D smooth basis function. We applied Karhunen-Loeve transform to a training set of smooth background images and the optimal transform turns out to be very similar to 2D DCT bases. Therefore, we used a set of low frequency two-dimensional DCT basis functions, since they have been shown to be very efficient for image representation [7].

B. Soft Edge Smoothness Prior to Alpha Channel Super Resolution:

This paper by the author Shengyang Dai proposed that effective image prior is necessary for image super resolution, due to its severely under-determined nature. Although the edge smoothness prior can be effective, it is generally difficult to have analytical forms to evaluate the edge smoothness, especially for soft edges that exhibit gradual intensity transitions. This paper finds the connection between the soft edge smoothness and a soft cut metric on an image grid by generalizing the Geocuts method, and proves that the soft edge smoothness measure approximates the average length of all level lines in an intensity image.

The new finding not only leads to an analytical characterization of the soft edge smoothness prior but also gives an intuitive geometric explanation. Regularizing the super resolution problem by this new form of prior can simultaneously minimize the length of all level lines, and thus resulting in visually appealing results. In addition, this paper presents a novel combination of this soft edge smoothness prior and the alpha matting technique for colour image super resolution, by normalizing edge segments with their alpha channel description, to achieve a unified treatment of edges with different contrast and scale.

Low resolution images are generated by smoothing and down-sampling target scenes with low quality image sensors. The task of recovering the original high resolution (HR) image from a single low resolution (LR) input is an inverse problem of this generation procedure. One criterion for solving this inverse problem is to minimize the reconstruction error. In other words, the result which can produce the same low-resolution image as the input one is preferred. Back-projection is proposed to minimize the reconstruction error efficiently by an iterative algorithm. However, since a lot of information is lost in the generation process, this problem is severely underdetermined. There might be multiple solutions to minimize this error, even for multiple LR input images. To overcome this difficulty, image priors need to be incorporated for regularizing the inverse problem [6].

C. Super-Resolution Image Reconstruction:

This paper by the author Sung Cheol Park proposed that the SR image reconstruction is proved to be useful in many practical cases where multiple frames of the same scene can be obtained, including medical imaging, satellite imaging, and video applications. The basic premise for increasing the spatial resolution in SR techniques is the availability of multiple LR images captured from the same scene. Registration is a very important step to the success of the SR image reconstruction.SR image reconstruction is one of the most spotlighted research areas because it can overcome the inherent resolution limitation of the imaging system and improve the performance of most digital image processing applications. Robustness and flexibility in modelling noise characteristics and a priori knowledge about the solution are the major advantages of the stochastic SR approach.

The term “SR image reconstruction” refers to a signal processing approach toward resolution enhancement because the term “super” in “super resolution” represents very well the characteristics of the technique overcoming the inherent resolution limitation of LR imaging systems. The major advantage of the signal processing approach is that it may cost less and the existing LR imaging systems can be still utilized. The SR image reconstruction is proved to be useful in many practical cases where multiple frames of the same scene can be obtained, including medical imaging, satellite imaging, and video applications. One application is to reconstruct a higher quality digital image from LR images obtained with an inexpensive LR camera/camcorder for printing or frame freeze purposes.

The SR technique is also useful in medical imaging such as computed tomography (CT) and magnetic resonance imaging (MRI) since the acquisition of multiple images is possible while the resolution quality is limited. In satellite imaging applications such as remote sensing and LANDSAT, several images of the same area are usually provided, and the SR technique to improve the resolution of the target can be considered. Another application is a conversion from an NTSC video signal to an HDTV signal since there is a clear and present need to display an SDTV signal on the HDTV without visual artifacts [3].

D. Image Up sampling via Imposed Edge Statistics:

This paper by the author Raanan Fattal proposed that a new method for up sampling images which can generate sharp edges with reduced input resolution grid-related artifacts. The method is based on a statistical edge dependency relating certain edge features of two different resolutions, which is generically exhibited by real-world images. While other solutions assume some form of smoothness, we rely on this distinctive edge dependency as our prior knowledge in order to increase image resolution.

In addition to this relation, we require that intensities are conserved; the output image must be identical to the input image when down sampled to the original resolution. Altogether the method consists of solving a constrained optimization problem, attempting to impose the correct edge relation and conserve local intensities with respect to the low-resolution input image. Results demonstrate the visual importance of having such edge features properly matched, and the capability of the method to produce images in which sharp edges are successfully reconstructed.

There is a unique dependency between image derivatives at different resolutions, as exhibited by real-world images; pixel differences at higher resolutions depend on their distance from an edge, the spatial distribution of that edge and the total intensity jump across it, all estimated in low-resolution. Using this non-trivial relation, a new method for up sampling images is devised. The solution consists in promoting the predicted intensity differences in the up sampled image given the edge parameters observed at the low-resolution input. This is done while deducing absolute intensities from an ‘intensity conservation’ constraint that requires the total intensity in the low and high resolutions to be the same. Given a low-resolution image plus this additional parametric statistical information, sharp edges are retrieved while typical artifacts associated with up sampling are minimal.

By real-world images, we refer to scenes seen with the naked eye or more precisely, scenes captured by a photographic device (e.g. camera). This includes indoor and outdoor photos not enlarged by any digital means. In the case of desktop publishing, raw images are resized, on a regular basis, to new dimensions to fit designated areas in documents. Low-resolution video frames from surveillance cameras are enlarged to ease the inspection of their contents. As well as the recent popularity of HDTVs brings out the need for resolution enhancement of NTSC and PAL video recordings.

In 3D graphics, these interpolations are used to map image textures onto objects' surface. While satisfactory down sampled images are obtained by a proper linear pre-filtering, this is not the case for up sampling. Up sampled images usually lack small-scale texture-related features and moreover, sharp edges become blurry, original pixel grids remain noticeable (often called the 'jaggies' artifact), and in some cases, ringing appears near sudden transitions in intensity. Formally speaking, up sampling involves determining far more pixel intensities than the number given. This makes up sampling a challenging problem and one that is highly sensitive to the additional assumptions or information needed to establish its well-posedness. Indeed, different up sampling techniques correspond to different assumptions about the nature of the up sampled image [5].

E. *Single Image Super-Resolution:*

This paper by the author Abhishek Arora proposed that:

1) In first approach, HR patches are constructed using a sparse representation of the corresponding LR patch in a compact dictionary.

2) In the second approach, exploit the repetition of local visual content within and across different scales of the given LR image.

Image super-resolution is the task of obtaining a high-resolution (HR) image of a scene given low-resolution (LR) image(s) of the scene. High resolution means high pixel density, also referred to as high-definition (HD). An HR image brings out details that would be blocked out in an LR image. More pixels in the same area implies a higher sampling frequency thereby offering more details. Bicubic interpolation is frequently used to increase the number of pixels in an image. However, it cannot recover original high frequency details of the scene if the scene is not sampled at a rate higher than the Nyquist frequency.

If the scene is not sampled at a rate higher than Nyquist frequency, then high frequency details are lost and cannot be recovered from individual single images. Most of the super resolution approaches presented in literature can be classified as:

1) Classical multi-image based super resolution: Classical multi-image super resolution refers to obtaining a higher resolution image of a scene from a sequence of lower resolution images of the same scene. Each low-resolution image places a set of linear constraints on the pixels of required HR image leading to a determined linear set of equations with HR pixels as variables. Provided a set of LR images {L1, L2..., Ln} of a scene with sub-pixel misalignments, these images can be used to recover a higher resolution image of the scene. However, this approach is numerically limited to a small increase in resolution. The limitations of this approach can be overcome using example-based approach.

2) Example-based super resolution: Example based super resolution refers to learning LR/HR patch correspondence from known LR/HR image pairs in a database, which provides a good prior on the predicted HR patch for a given LR patch. This technique is not guaranteed to recover the actual high frequency details [8].

F. *Joint dictionary learning for example-based image super-resolution:*

Mohsen Joneidi proposed: A new joint dictionary learning method for example-based image super resolution (SR), using sparse representation. The low-resolution (LR) dictionary is trained from a set of LR sample image patches. Using the sparse representation coefficients of these LR patches over the LR dictionary, the high-resolution (HR) dictionary is trained by minimizing their construction error of HR sample patches. The error criterion used here is the mean square error. In this way, the HR patches have the same sparse representation over HR dictionary as the LR patches over the LR dictionary, and at the same time, these sparse representations can well reconstruct the HR patches.

Prior to SR methods, the usual way to increase the resolution of images was to use simple interpolation-based methods such as bilinear, bicubic and more recently the resampling methods. However, all these methods suffer from blurring high-frequency details of the image especially for large upscaling factors (the amount by which the resolution of the image is increased in each dimension). Thus, over the last few years, many SR algorithms have been proposed. These methods can be classified into two categories: multi-image SR, and single-image SR. In the conventional SR problem, multiple images of the same scene with subpixel motion are required to generate the HR image. However, the performance of these SR methods is only acceptable for small upscaling factors (usually smaller than 2). As the upscaling factor increases, the SR problem becomes severely ill-conditioned and many LR images are needed to recover the HR image with acceptable quality.

The usual way to increase the resolution of images was to use simple interpolation-based methods such as bilinear, bicubic and more recently the resampling method. However, all these methods suffer from blurring high-frequency details of the image, especially for large upscaling factors the amount by which the resolution of the image is increased in each dimension. Recently, joint and coupled learning methods are utilized for efficient modelling of correlated sparsity structures. However joint learning methods and the coupled learning methods still do not guarantee that the sparse representation of HR image patches over the HR dictionary is the same as the sparse representation of LR patches over LR dictionary. To address this problem, a direct way to train the dictionaries that enforces the same sparse representation for LR and HR patches is proposed. Moreover, since the HR dictionary is trained by minimizing the final error in reconstruction of HR patches [9].

3. CONCLUSION

Sparsity leads to a powerful prior to the ill-posed problem of single image super resolution. Cross channel information and colour constraints lead to regularizing the optimization problem for boosting SR performance. Under different scaling factors, different noise levels, different dictionary sizes we have discussed certain methods that outperform the state of the art. Expedite the sparse coding problem using neural networks. Introduce other objective measurements rather than MSE for quality assessment or in the objective function. Apply other cross channel constraints or colour constraint that can improve super resolution performance. The sparsity based super-resolution is extended to multiple colour channels. We demonstrate that by using colour information and cross channel constraints, a significant improvement over single (luminance) channel sparsity based SR methods can be achieved. Edge similarities among colour bands are exploited as cross channel correlation constraints. These additional constraints lead to new optimization problems in the sparse coding.

REFERENCES

[1] Hojjat S. Mousavi and Vishal Monga," Sparsity based Colour Image Super Resolution via Exploiting Cross Channel Constraints", Transactions on Image Processing, April 2017.

[2] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar," Fast and robust multiframe super resolution," IEEE Trans. on Image Processing, vol. 13, no. 10, pp. 1327-1344, 2004.

[3] S. C. Park, M. K. Park, and M. G. Kang," Super-resolution image reconstruction: a technical overview," Signal Processing Magazine, IEEE, vol. 20, no. 3, pp. 21-36, 2003.

[4] M. F. Tappen, B. C. Russell, and W. T. Freeman," Exploiting the sparse derivative prior for super-resolution and image demosicing," in IEEE Workshop on Statistical and Computational Theories of Vision, 2003.

[5] R. Fattal," Image up sampling via imposed edge statistics" in ACM Trans. Graph. (TOG), vol. 26, no. 95.

[6] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong" Soft edge smoothness prior to alpha channel super resolution" in Proc. IEEE Conf. Computer Vision Pattern Recognition, 2007, pp.1-8.

[7] S. Minaee and Y. Wang, "Screen content image segmentation using least absolute deviation fitting," in Proc. IEEE Conf. on Image Processing, 2015, pp. 3295-3299.

[8] D. Glasner, S. Bagon, and M. Irani," Super-resolution from a single image," in Proc. IEEE Conf. on Computer Vision, 2009, pp. 349-356.

[9] Mojtaba Sahraee-Ardakan, Mohsen Joneidi,"Joint Dictionary Learning for Example-Based Image Super-Resolution" Computer Vision and Pattern Recognition, 2017.