Super-resolving a Single Intensity/Range Image via Non-local Means and Sparse Representation

Srimanta Mandal,* Arnav Bhavsar and Anil K. Sao
School of Computing and Electrical Engineering
Indian Institute of Technology Mandi, HP, India
srimanta_mandal@students.iitmandi.ac.in, arnav@iitmandi.ac.in, anil@iitmandi.ac.in

ABSTRACT

We propose an example-based super-resolution (SR) framework, which uses a single input image and, unlike most of the SR methods does not need an external high resolution (HR) dataset. Our SR approach is based in sparse representation framework, which depends on a dictionary, learned from the given test image across different scales. In addition, our sparse coding focuses on the detail information of the image patches. Furthermore, in the above process we have considered non-local combination of similar patches in the input image, which assist us to improve the quality of the SR result. We demonstrate the effectiveness of our approach for intensity images as well as range images. Contemplating the importance of edges in images of both these modalities, we have added an edge preserving constraint that will maintain the continuity of edge related information to the input low resolution image. We investigate the performance of our approach by rigorous experimental analysis and it shows to perform better than some state-of-the-art SR approaches.

Keywords

Super-Resolution, Non-local similarity, Sparse domain, Edge preservation.

1. INTRODUCTION

Super-Resolution (SR) is a process of increasing resolution of images. Recently, availability of high resolution (HR) sensors for intensity images at lower costs may pose a question for the need of super-resolution algorithms for this type of images. Still, there are some physical aspects like distance, bandwidth requirement, storage etc. are present in current scenario, which prompts us to super-resolve intensity images. On the other hand high quality range image[1] finds its application in computer vision, multimedia etc., and there is a requirement of enhancing resolution of range images, captured by low cost range cameras like ToF cameras, Kinect cameras etc [7, 9].

*Corresponding author

[1] In this paper, depth and range both words have been used interchangeably.

ICVGIP '14, December 14-18, 2014, Bangalore, India

http://dx.doi.org/10.1145/2683483.2683541.

Interpolation techniques [12, 13] can approximate the increase in resolution of an image but they often fail to preserve subtle details. As a result, SR techniques have evolved so as to achieve the goal of preserving detail information in case of intensity image. In recent years, single image SR methods are substituting multi-image SR approaches as the former is more useful in practical scenarios and has also shown better performance. Though, single image SR method, by definition doesn't require multiple images of the target scene, the process does require some HR example images [4, 14, 18, 21, 22]. The relationship learned from the HR-LR (high resolution-low resolution) patch pair from these example images form the basis of example based single image SR.

In case of range images, most of the SR approaches borrow the prominent edge and related information from a HR color image of the target scene [7, 10]. This is because, the discontinuities present in the range images concur with those in color image. But some scenarios like transmission of depth information, 3-D modeling etc. may force us to reconstruct dense depth information using only a single LR range image. It can be addressed by example based SR methods in a similar fashion as is done in case of intensity image SR. The example based methods [1] devoid of registration which might be required for aligning range and color images in case of acquisition made using uncalibrated camera for both types of images.

The performance of the example based SR approach depends on the relationship, that has been learned from HR-LR patch pairs. This requires a large number of example images to be collected for training purpose. Yang et al. [21] and Zeyde et al. [22] exploit the same principal to learn HR-LR dictionary pair in sparse domain framework. The approaches stated in references [4, 14] have learned multiple HR sub-dictionaries to employ the subtle details and implant them adaptively in the LR patches. But an important shortcoming of all these methods is relying on predefined dictionary(s) for all kinds of image variations. In addition, the absence of similar patches to the test patch in the trained dictionary may degrade the SR outcomes. Furthermore, learning the relationship between HR-LR patch pairs from a large number of images increase the computational time.

These issues have been addressed in this work. We propose a SR framework which doesn't use any image other than the input LR image. Indeed, some works have been reported in the literature in the same direction [8, 20]. Glasner et al. [8] proposed a method which utilizes the patch redundancies present in the same scale as well as across different scales. The principal idea behind this method is the presence of similar patches in the image at different scales. Whenever, a patch similar to the test patch has been found in a down-scaled version of the image, the parent patch of the same has been copied to the appropriate location of the HR image grid.

The work has been extended by incorporating a group sparsity constraint in reference [20]. Unlike these methods, we extract patches of same dimension from the LR test image across different scales and cluster them according to the detail information present in the patches. We then learn compact sub-dictionaries from these clusters. Moreover, we focus on using the detail information for SR. We extract such detail information in an elegant non-local mean approach and this is followed up by sparse coding the same with the help of learned sub-dictionaries. In addition, we have added an effective edge preserving constraint, which has been proposed in the work [14]. This constraint will compel the SR outcome to follow the similar edge like information to the LR test image. Thus, the effectiveness of our approach is due to the sparse representation involving an edge preservation and adaptive sub-dictionaries, which encodes the information about patch details, computed using non-local means. In contrast with most of the SR approaches, we have demonstrated the effectiveness of our SR approach in case of intensity images as well as range images.

Thus, our contributions can be summarized as: 1) Our approach obviates the need of any extra image in SR. Though some works have been published towards similar direction for intensity image SR, but this approach is completely new in case of range image SR. Moreover, we have learned multiple dictionaries from the patches extracted from the test image to work in sparse environment. 2) We super-resolve perceptually important detail information extracted from the patches by using non-local mean operation. 3) The edge preserving constraint plays a crucial role in preserving edge related information in the SR outcome. 4) We demonstrate the improvement of SR in case of intensity as well as range images.

The the rest of the paper is organized as follows: Section 2 illustrates the background of SR in sparse domain. Section 3 discusses the proposed approach in greater details. The experimental results have been analyzed in section 4. Finally, the paper is concluded in section 5.

2. SR IN SPARSE DOMAIN

Mathematical formulation of the process of LR image formation provides us clues about getting the HR image back and one of the most popular mathematical model is

$$\mathbf{y} = \mathbf{DHx} + \boldsymbol{\nu}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the observed LR image which has been generated by blurring ($\mathbf{H} \in \mathbb{R}^{n \times n}$) and down-sampling ($\mathbf{D} \in \mathbb{R}^{m \times n}$) the HR scene $\mathbf{x} \in \mathbb{R}^n$. In this case $n > m$ and $\boldsymbol{\nu}$ is the noise component. It is now clear that SR is an inverse problem of finding $\mathbf{x}$ from its LR observation $\mathbf{y}$. Since $n > m$, the problem of finding an $\mathbf{x}$ from $\mathbf{y}$ became under-determined as there can be many $\mathbf{x}$ to produce the same $\mathbf{y}$. Thus regularization approaches have been incorporated into the problem such as Tikhonov regularization [23], Total Variation (TV) regularization etc. [15]. These methods regularize the SR problem but fail to protect subtle details like edges, textures etc. in the SR outcome. Thus, a lot of regularization methods have been unfolded and one of the recently proposed approach is sparsity regularization, where image is represented in sparse domain.

Here, an image is represented as a linear combination of few columns of a dictionary matrix ($\mathbf{A}$), thus $\mathbf{x} = \mathbf{Ac}$. $\mathbf{c}$ being the sparse coefficient vector, plays an important role in finding the suitable atoms from $\mathbf{A}$. Thus the goal is to find $\mathbf{c}$ and can be estimated by solving the following optimization problem [6]:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \left\{ \|\mathbf{y} - \mathbf{DHAc}\|_2^2 + \lambda \|\mathbf{c}\|_1 \right\}, \tag{2}$$

where $\|\cdot\|_1$: finds the l_1-norm of a vector and Lagrangian multiplier λ weights between data term $\|\mathbf{y} - \mathbf{DHAc}\|_2^2$ and sparsity term $\|\mathbf{c}\|_1$. Here, l_1-norm has been used to enforce sparsity on $\mathbf{c}$ instead of l_0-norm as it's computation asks for combinatorial search and is NP-hard in nature. It has been found that l_1-norm minimization is the closest convex optimization of l_0-norm minimization [5]. From eqn. (2), one can observe that the computation of $\hat{\mathbf{c}}$ involves requirement of $\mathbf{D}, \mathbf{H}$ and $\mathbf{A}$. $\mathbf{D}$ and $\mathbf{H}$ are typically assumed to be known and the dictionary matrix $\mathbf{A}$ is the most important source of information in SR. The process of finding the dictionary $\mathbf{A}$ is discussed in the following section.

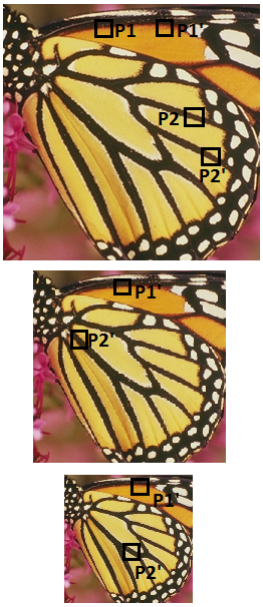

Figure 1: Illustration of patch similarity for intensity images in same scale and across different scales.

3. PROPOSED APPROACH

The proposed work super-resolve the input LR image without assistance of other HR image. It involves learning the coarse to fine information of patches from the LR test image and keeping in the form of sub-dictionaries and has been discussed in the subsection 3.1. Keeping in mind the significance of detail information, we extract and restore them by the help of the learned sub-dictionaries and non-local similar patches. This reconstruction process has been discussed in subsection 3.2.

3.1 Learning Dictionary:

In current scenario, the LR test image ($\mathbf{y}$) is the only source of information. To exploit the available information in appropriate way, we have to investigate similar information in image pyramid by up/down-scaling the input image. One can examine such pyramids for intensity as well as range images in Figures 1 and 2 respectively. It can be observed that the patches similar to P1 for intensity image can be found in the same scale and across different scales and those are found to be P1'. Same is true for the range image also. In fact, the region present at the same distance from the range camera will have same intensity. Thus, there will be plenty of similar patches available for range images. These similar patches across scales will provide coarse to fine information.

To grab such information we interpolate the LR test image to the HR image grid and extract patches of size $\sqrt{p} \times \sqrt{p}$ from it. We down-sample the interpolated image in three levels by s^k factors to

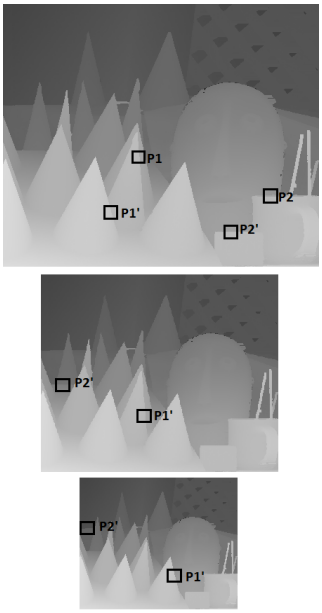

Figure 2: Illustration of patch similarity for range images in same scale and across different scales.

complete the image pyramid. Again the procedure of patch extraction is followed for down-sampled versions of the interpolated image. This operation will allow us to have more patches for training. Thus, we have patches of same dimension from different resolutions. All the extracted patches have been clustered using K-means clustering algorithm depending on their detail information[2]. This process will group the raw patches with similar detail information across different scales and same scale together into a cluster.

Due to clustering of similar patches extracted across different scales, we may get fine information related to coarse information for the target patch. These information will assist us in reconstructing the HR image. To give importance to perceptually important detail component like edges, corners etc. we extract the detail information by subtracting the mean component from each of the clusters and analyze them based on their principal components to achieve compact sub-dictionaries. Thus, we have some sub-dictionaries $\mathbf{A}_k$ and there representative centroids $\boldsymbol{\mu}_k$. It has to be noted that the centroids are achieved from the K-means clustering and are related to the detail information of the corresponding cluster. For more discussion on how the principal components of each cluster have been analyzed one may refer to the article [4].

3.2 HR image reconstruction:

Our reconstruction starts with interpolating the LR test image to the HR grid. This interpolated image $\hat{\mathbf{x}}$ will work as an initial approximation of the unknown HR image. This image lacks detail information like textures, edges, corners etc. and these need to be restored, which is very important from human perception point of view. This argument is invalid in case of range image, as these are not directly perceived by human beings and are used for some applications only. But if we examine any range image (say Fig. 2), we will be able to find that the most important component of range images are the detail information like edges, corners etc. and if we enhance those properly, a good SR outcome is expected.

[2]Here, the 'detail information' is computed by subtracting low pass filtered patch from its original version.

Thus, we focus to reconstruct detail information by restoring in elegant sparse coding environment. Unlike traditional methods of computing detail information, we subtract the weighted average of non-local similar patches from the target patch in the image. The reason being that the non-local mean will contain all the variations of the smooth component and thus, if we subtract this from the test patch an appropriate representation of the detail information can be achieved.

Here, the patches are extracted from the image by $\mathbf{x}_i = \mathbf{P}_i\hat{\mathbf{x}}$, where $\mathbf{P}_i$ is assumed to be a patch extractor matrix. There will be several similar patches as illustrated in Figures 1&2, and their weighted average will be similar to the test patch but the average won't be exactly same to the test patch. That difference is the detail information which is missing in test patch. Let $\mathbf{x}_{i,m}$ be the index of similar patches of the test patch $\mathbf{x}_i$ and are kept in the set $\boldsymbol{\zeta}_i$. Thus, the non-local mean can be calculated as:

$$\overline{\mathbf{x}}_i = \sum_{m\in\zeta_i} w_{i,m}\mathbf{x}_{i,m}, \tag{3}$$

where the weight $w_{i,m}$ depends on the similarity of patches and it lies within a range of 0 to 1. This weight is measured as a decreasing function of weighted Euclidean distance [2]

$$w_{i,m} = \frac{1}{z}e^{-\frac{\|\mathbf{x}_i-\mathbf{x}_{i,m}\|_2^2}{h}}, \tag{4}$$

z is the normalizing constant and h controls the decay of the exponential function which decreases with the weight of Euclidean distance. The usage of Euclidean distance is logical in this case as it preserves the order of similarity between pixels in presence of additive noise.

Once, we have the non-local mean, the differences between the test patch and the non-local mean can be computed by

$$\mathbf{d}_{\mathbf{x}_i} = |\mathbf{x}_i - \overline{\mathbf{x}}_i|. \tag{5}$$

This $\mathbf{d}_{\mathbf{x}_i}$ is the detail information of the test patch, which will be sparse coded with the assistance of learned sub-dictionaries. [3] Now, the task is to select a sub-dictionary $\mathbf{A}_k$ for the test patch $\mathbf{x}_i$. Since, $\boldsymbol{\mu}_k$ are the representatives of corresponding dictionaries, the selection can be done based on simple Euclidean distance between $\mathbf{d}_{\mathbf{x}_i}$ and $\boldsymbol{\mu}_k$

$$k_i = \arg\min_k \ \|\mathbf{d}_{\mathbf{x}_i} - \boldsymbol{\mu}_k\|_2, \tag{6}$$

where k_i be the index of the selected dictionary $\mathbf{A}_k$ for the patch $\mathbf{x}_i$.

Next, sparse coefficient for the difference component $\mathbf{d}_{\mathbf{x}_i}$ is computed by solving the following cost function

$$\hat{\mathbf{c}}_{d_i} = \arg\min_{\mathbf{c}_{d_i}} \left\{\|\mathbf{d}_{\mathbf{x}_i} - \mathbf{A}\mathbf{c}_{d_i}\|_2^2 + \lambda\|\mathbf{c}_{d_i}\|_1\right\}, \tag{7}$$

and is solved by iterative thresholding algorithm as proposed in [3]. Since, we are selecting a particular sub-dictionary from all the sub-dictionaries, the computed coefficient vector is happened to be highly sparse. This coefficient $\hat{\mathbf{c}}_{d_i}$ allow us to get back the reconstructed detail component of the patch by $\hat{\mathbf{d}}_{\mathbf{x}_i} = \mathbf{A}_k\hat{\mathbf{c}}_{d_i}$. The $\hat{\mathbf{d}}_{\mathbf{x}_i}$ lacks the slow varying component and can be compensated by adding the non-local mean $\overline{\mathbf{x}}_i$ to produce the super-resolved raw patch

$$\hat{\mathbf{x}}_i = \hat{\mathbf{d}}_{\mathbf{x}_i} + \overline{\mathbf{x}}_i. \tag{8}$$

[3]Note that the detail information in a sub-dictionary is computed by subtracting the cluster-mean. This is slightly different from the computation of $\mathbf{d}_{\mathbf{x}_i}$ which involves subtracting the non-local-mean. This difference does not affect the sparse coding significantly, but yields a considerable computational advantage in dictionary learning.

Once, all the patches are reconstructed from its LR version, the entire image can be reconstructed back by

$$\hat{\mathbf{x}} \approx \left(\sum_{i=1}^{L} \mathbf{P}_i^T \mathbf{P}_i \right)^{-1} \sum_{i=1}^{L} \left(\mathbf{P}_i^T \hat{\mathbf{x}}_i \right), \tag{9}$$

where L denotes the total number of patches. The eqn. (9) demonstrates that all the reconstructed patches are kept in the corresponding position of HR image grid as they were in LR image with the averaged overlapping portions. The reconstructed image should look similar to the input LR image and to make sure this fact, we minimize the following cost function

$$\hat{\hat{\mathbf{x}}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{DH}\hat{\mathbf{x}}\|_2^2. \tag{10}$$

In order to achieve better localization, the eqn. (10) is further regularized by an effective edge preserving constraint as proposed in [14].

Here *edginess* feature has been considered to preserve and is known to perform better than conventional gradient [16]. This is generally computed by applying a smoothing operator along one direction and its derivative operator along orthogonal direction. Considering $\mathbf{E}_g$ be the operator responsible for extracting the gradient magnitude $\mathbf{e}_g = \sqrt{\mathbf{e}_0^2 + \mathbf{e}_{90}^2}$, where $\mathbf{e}_0$ is the vertical edge evidence and $\mathbf{e}_{90}$ is the horizontal edge evidence. Thus the eqn. (10) can be rewritten with the edge preserving constraint as:

$$\hat{\hat{\mathbf{x}}} = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{DH}\hat{\mathbf{x}}\|_2^2 + \beta \|\mathbf{E}_g\{\mathbf{y}\} - \mathbf{E}_g\{\mathbf{DH}\hat{\mathbf{x}}\}\|_2^2 \right\} \tag{11}$$

The edge preserving constraint minimizes the difference between edginess information of LR image and that of the down-sampled version of the reconstructed image. As a result, it will help to preserve perceptually significant discontinuities present in intensity image as well as range image. Finally, we have the estimated HR image $\hat{\hat{\mathbf{x}}}$ and the procedure (Eqns. (3) to (11)) needs to be iterated until convergence for better results. Within each iteration the sub-dictionary will be selected adaptively.

The pseudo code of the proposed approach is summarized in **Algorithm 1**, where $Shrink$ operator is used for solving eqn. (7) by using iterative shrinkage algorithm as proposed in [3] with the help of a predefined parameter λ and $\mathbf{E}$ can be assumed to be the edge extraction matrix similar to the operator $\mathbf{E}_g$. It has to be noted that, the most inner loop will run for L times, which is the number of patches. Second inner loop will check for convergence and the outer loop is optional and can be used to update the sub-dictionaries. Thus, within every iteration the dictionary selection will be updated adaptively and once the inner loops are completed, the dictionary will be retrained. The outcome of this algorithm is analyzed and compared with the state-of-the-art approaches in the next section.

Algorithm 1: Single Image SR

Data: LR image $\mathbf{y}$
Result: HR image $\mathbf{x}$

Initialization:
Set initial approximation $\hat{\mathbf{x}} = (\mathbf{y}) \uparrow_d$
Set the regularization parameters λ, β and γ.
Set error threshold ϵ.
Main Iteration:
for $k = 1$ **to** K **do**
  Extract patches $\mathbf{x}_i$ by $\mathbf{x}_i = \mathbf{P}_i\hat{\mathbf{x}}$ from $(\hat{\mathbf{x}}) \downarrow_d$ for $d = 1$ to s.
  Apply K-means clustering on all $\mathbf{x}_i$ and get $\boldsymbol{\mu}_k$.
  Apply PCA to each cluster to learn several $\mathbf{A}_k$.
  for $j = 1$ **to** N **do**
    for $i = 1$ **to** L **do**
      Search for similar patches $\mathbf{x}_{i,m}$.
      Calculate the non-local mean by $\overline{\mathbf{x}}_i = \sum_{m \in \zeta_i} w_{i,m}\mathbf{x}_{i,m}$.
      Compute $\mathbf{d}_{\mathbf{x}_i} = |\mathbf{x}_i - \overline{\mathbf{x}}_i|$.
      Select a particular $\mathbf{A}_k$ for $\mathbf{d}_{\mathbf{x}_i}$ based on eqn. (6).
      Compute the coefficient vector $\hat{\mathbf{c}}_{d_i} = Shrink\left(\mathbf{A}_k^T \mathbf{d}_{\mathbf{x}_i}\right)_\lambda$.
      Restore the patch $\hat{\mathbf{x}}_i = \mathbf{A}_k\hat{\mathbf{c}}_{d_i} + \overline{\mathbf{x}}_i$.
    end
    Achieve the full image by eqn. (9).
    $\hat{\mathbf{x}}^{j+1} = \hat{\mathbf{x}}^j + \gamma(\mathbf{DH})^T\left(\mathbf{y} - \mathbf{DH}\hat{\mathbf{x}}^j\right) + \beta(\mathbf{EDH})^T\left(\mathbf{Ey} - \mathbf{EDH}\hat{\mathbf{x}}^j\right)$
    if $\|\hat{\mathbf{x}}^{j+1} - \hat{\mathbf{x}}^j\|_2^2 < \epsilon$ **then**
      break;
    else
      continue;
    end
  end
end

4. EXPERIMENTAL RESULTS

We now discuss some experimental results of our proposed approach. For better clarification, we have divided this section into two subsections one for SR results of the intensity images and other for the results of range image.

4.1 SR results of intensity images:

Here we have used some standard intensity images of dimension 256×256. According to the LR image formation model (1), these HR images are first blurred by a 7×7 Gaussian kernel with standard deviation 1.6 and down-sampled by factor 3. The down-sampling has been done by leaving 3 pixels in both horizontal and vertical directions. Thus, the LR images to be super-resolved is of dimension 56×56, which is due to the round-off operation on $256/3$. This LR image is interpolated by bi-cubic interpolation technique, which will act as an initial approximation of the unknown HR image. The patch size we have selected in our experiments is 6×6. The same dimensional patches have been extracted from the down-sampled version of the bi-cubic interpolated image. Here, we consider $(0.8)^{2n}$ as s^k, where $n = 1, 2, 3$. In this case, the number of patches for training are on the order of 100,000. The value of K in K-means clustering is chosen as 68. We require such a high number in training as lesser number may wash out the differences among the clusters and on the other hand too large number makes each cluster less informative. Thus, we classify the set of image patches into 68 clusters and those clusters with very few patches are being merged with the neighbor classes.

In searching for similar patches, the weight has been assigned depending on the similarity to the target patch. This eliminates the requirement of a threshold for similarity measurement. In the algorithm the values of λ, β and γ are empirically chosen as 0.08, 0.01 and 7 respectively. The results of the algorithm are being compared with some best performing state-of-the-art single image SR approaches [4, 21, 22] qualitatively and quantitatively by their publicly available codes [4, 5]. For quantitative comparison we have used

[4] The source code of [21, 22] is available at http://www.cs.technion.ac.il/~elad/Various/Single_Image_SR.zip

[5] The source code of [4] is available at http://www4.comp.polyu.edu.hk/~cslzhang/ASDS_data/TIP_ASDS_IR.zip

PSNR and SSIM [19], where PSNR measure the quality in error perspective and SSIM checks for similarity with the original image and is related to the human visual system. In case of color images, we consider perceptually important luminance component for SR and latter the chromatic component is added back to produce the final HR color image.

We super-resolve the images by up-sampling factor 3 and the results can be observed in Fig. 3 and Fig. 4 for *Butterfly* and *Plant* image. For both figures, the top left image is the LR image zoomed up to the same scale of HR image, top middle image is the SR result of the approach [21] and top right image is the result of [22]. In the bottom row, left one is the result of [4] and the result of our approach is placed in the middle position and the place of original image is bottom right.

One can observe that the results of the approaches [21, 22] are smoother in the areas of discontinuities in comparison to our approach. The performance of the approach [4] is very much closer to ours. In fact, visually there is little difference between these. It has to be noted that all these approaches consider external HR images for training dictionary, whereas our approach doesn't need any external images. Still, some differences can be found quantitatively in the Table. 1, where the comparison among the approaches are shown in terms of PSNR and SSIM. One can note the improvements of our approach in comparison to state-of-the-art approaches.

Table 1: Results of SR for Intensity Images (↑ 3)

Images	Metrics	Bi-cubic	Raw Patch [21]	Scale Up [22]	ASDS [4]	Proposed Approach
Baboon	PSNR	19.72	21.60	**21.61**	20.70	20.71
	SSIM	0.3417	0.4363	0.4299	0.4936	**0.4951**
Barbara	PSNR	22.91	23.51	23.56	24.36	**24.38**
	SSIM	0.6155	0.6440	0.6451	0.7307	**0.7314**
Bike	PSNR	20.80	21.59	21.54	24.02	**24.26**
	SSIM	0.5756	0.6409	0.6348	0.7733	**0.7842**
Butterfly	PSNR	20.78	21.53	21.52	26.05	**27.00**
	SSIM	0.7173	0.7747	0.7801	0.8703	**0.8959**
Cameraman	PSNR	21.69	22.12	22.13	24.66	**24.93**
	SSIM	0.7025	0.7452	0.7490	0.8079	**0.8185**
Girl	PSNR	29.82	29.89	29.92	33.46	**33.55**
	SSIM	0.7317	0.7424	0.7417	0.8228	**0.8238**
Hat	PSNR	27.20	27.89	27.97	30.47	**30.94**
	SSIM	0.7773	0.8177	0.8214	0.8552	**0.8654**
Parrot	PSNR	25.58	26.10	26.06	29.68	**29.95**
	SSIM	0.8256	0.8439	0.8473	0.9055	**0.9098**
Peepers	PSNR	22.99	23.82	23.91	27.81	**28.19**
	SSIM	0.7217	0.7631	0.7713	0.8529	**0.8598**
Plants	PSNR	27.76	28.29	28.29	32.84	**33.41**
	SSIM	0.7845	0.8163	0.8151	0.9006	**0.9092**

4.2 SR results of range images:

In this case, we have used range images from the standard Middlebury dataset [11, 17]. The sizes of HR images are of the order 500×400. The database contains images with some missing pixels i.e. black pixels. In order to avoid false quantitative results, we filled up those black pixels with its available left nearest neighbor. These filled up range images are then blurred by the same Gaussian kernel as is used in case of intensity images and down-sampled by factors 2 and 4. The sizes of the down-sampled LR images, which have to be super-resolved are of the order 250×200 and 125×100 respectively. In case of range image, the number of patches for training purpose is 250,000. Rest of the parameters for SR are kept same as in case of intensity image SR.

Table 2: Results (RMSE) of SR for Range Images (↑ 2)

Images	EB [1]	GIF [10]	ATGV [7]	Ours
Aloe	5.58	5.93	5.07	**2.87**
Baby	3.35	3.27	2.97	**1.66**
Cones	4.41	3.93	3.51	**2.20**
Plastic	3.01	2.32	2.22	**1.27**
Teddy	3.38	3.01	2.67	**1.71**
Tsukuba	9.92	15.62	11.20	**5.24**
Venus	1.94	2.72	1.84	**0.93**

Table 3: Results (RMSE) of SR for Range Images (↑ 4)

Images	EB [1]	GIF [10]	ATGV [7]	Ours
Aloe	7.46	6.30	5.76	**4.10**
Baby	4.49	3.55	3.36	**2.63**
Cones	5.90	4.39	4.00	**3.33**
Plastic	4.35	2.66	2.31	**2.09**
Teddy	5.20	3.32	3.08	**2.46**
Tsukuba	19.97	17.09	27.49	**9.37**
Venus	2.36	2.79	2.68	**1.40**

We super-resolve those LR images by up-sampling factors 2 and 4 and compare our results with the popular range image SR approaches, which involved requirement of color images [7, 10] and an example based range image SR approach [1], which doesn't use color images but it uses a HR database. For all these approaches we have used publicly available codes provided by the respective authors.

The results for up-sampling factor 2 can be observed for the *Aloe* image in Fig. 5. Here top left image is the input LR image, top middle stands for the result of the approach [10], the result of the approach [7] is placed in top right, bottom left shows the result of the example based approach [1], the bottom middle displays the result of our proposed approach and the bottom right is the original image. One can clearly examine the results and point out that our approach produces the best result in comparison to other approaches, which failed to preserve the edges. As can be seen in figure that edges are smeared for most of the approaches. This smearing of edges increase as the up-sampling factor increases and can be observed for factor 4 in Fig. 6 and 7. It has to be noted that the sharpness of the results of our approach has come down with increasing factor but still it is able to contain the strong edges intact, which is not the case with other approaches.

Now, we show some quantitative results of the SR approaches in terms of root mean square error (RMSE). Here, we do not consider SSIM in evaluating the image quality, because SSIM is related to HVS but the range images are not directly perceived by human beings. The RMSE values of the SR results for up-sampling factors 2 and 4 are kept in Tables 2 and 3 respectively. It can be observed, that for all the cases our proposed approach is performing better than the state-of-the-art approaches including color image based approaches.

Figure 3: Comparison of SR approaches for *Butterfly* image: Top left is the zoomed version of the LR image, top middle is the SR result of [21], top right is the SR result of [22]. Bottom left represents the SR results of [4], middle one is the result of the proposed approach and right bottom is the original image.

Figure 4: Comparison of SR approaches for *Plant* image: Top left is the zoomed version of the LR image, top middle is the SR result of [21], top right is the SR result of [22]. Bottom left represents the SR results of [4], middle one is the result of the proposed approach and right bottom is the original image.

5. CONCLUSION

We proposed a single image SR approach that doesn't need any external HR image, which is unlike to most other single image SR approaches. Our approach targeted for both intensity and range images. We mould the formulation of the approach in sparse representation framework, where we learn sub-dictionaries from the patches extracted from the input image across different scales to capture coarse to fine information. To achieve a good localization we have considered extracting the detail information based on elegant NL-mean formulation and have sparse-coded this information. The SR result is further improved by regularizing with an effective edge preserving constraint. Thus, the combination of non-local similarity, sparse representation and edge preservation plays the key role in our approach. We demonstrate the performance of our approach in case of range and intensity images. The experimental results show considerable improvement over the state-of-the-art approaches.

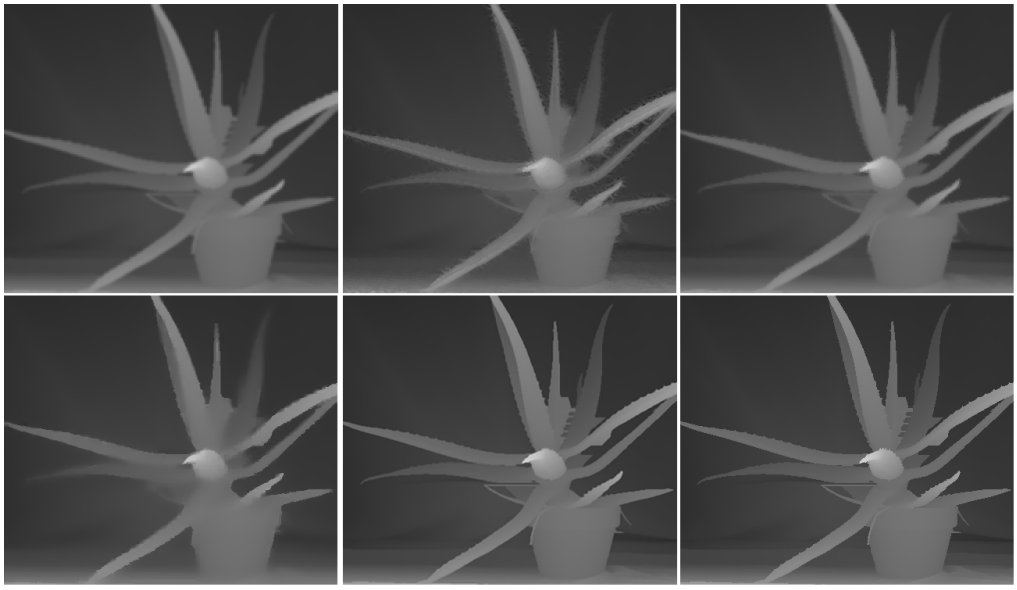

Figure 5: Comparison of SR approaches for *Aloe* image (scale-2): Top left is the zoomed version of the LR image, top middle is the SR result of [10], top right is the SR result of [7]. Bottom left represents the SR results of [1], middle one is the result of the proposed approach and right bottom is the original image.

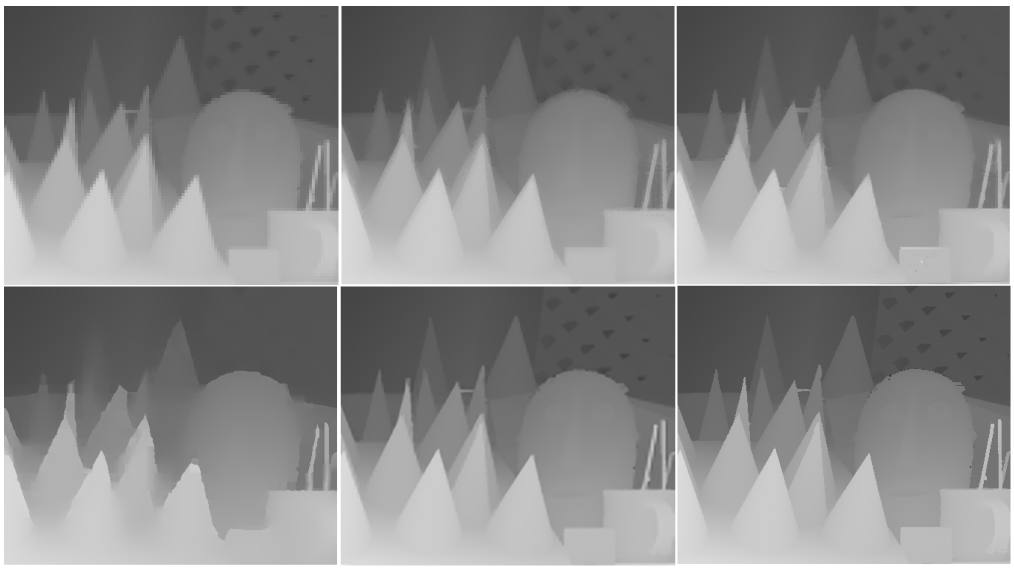

Figure 6: Comparison of SR approaches for *Cones* image (scale-4): Top left is the zoomed version of the LR image, top middle is the SR result of [10], top right is the SR result of [7]. Bottom left represents the SR results of [1], middle one is the result of the proposed approach and right bottom is the original image.

6. REFERENCES

[1] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow. Patch based synthesis for single depth image super-resolution. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ECCV'12, pages 71–84, Berlin, Heidelberg, 2012. Springer-Verlag.

[2] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, volume 2, pages 60–65 vol. 2, June 2005.

[3] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[4] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838 –1857, Jul. 2011.

[5] D. L. Donoho. For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 59(7):907–934, 2006.

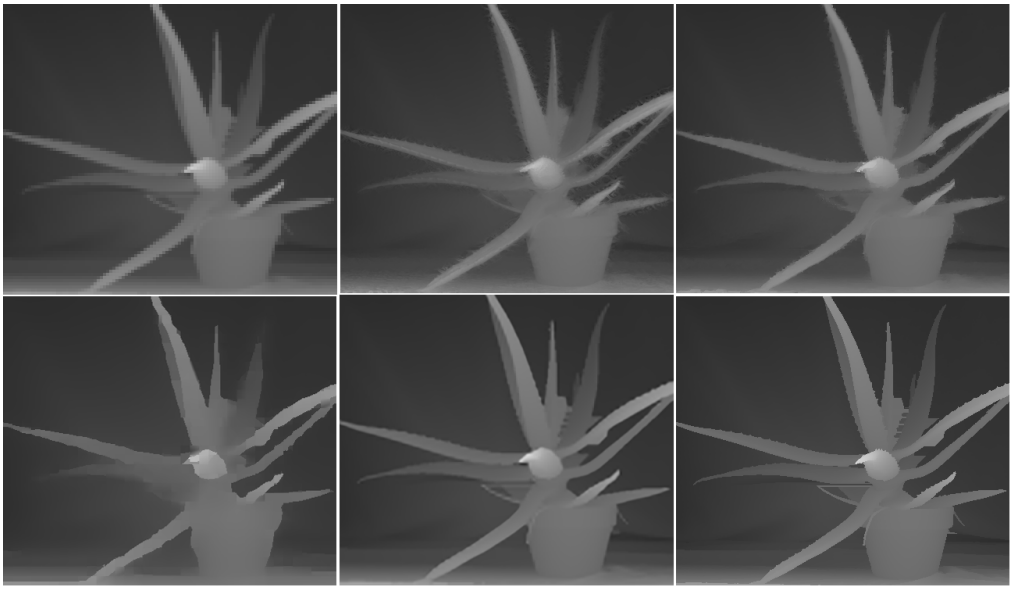

Figure 7: Comparison of SR approaches for *Aloe* image (scale-4): Top left is the zoomed version of the LR image, top middle is the SR result of [10], top right is the SR result of [7]. Bottom left represents the SR results of [1], middle one is the result of the proposed approach and right bottom is the original image.

[6] M. Elad. *Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[7] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *IEEE International Conference on Computer Vision (ICCV), 2013*, pages 993–1000, Dec 2013.

[8] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE 12th International Conference on Computer Vision, 2009*, pages 349–356, Sept 2009.

[9] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *Cybernetics, IEEE Transactions on*, 43(5):1318–1334, Oct 2013.

[10] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, June 2013.

[11] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1–8, June 2007.

[12] H. Hou and H. Andrews. Cubic splines for image interpolation and digital filtering. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(6):508–517, 1978.

[13] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6):1153–1160, 1981.

[14] S. Mandal and A. Sao. Edge preserving single image super resolution in sparse environment. In *20th IEEE International Conference on Image Processing (ICIP), 2013*, pages 967–971, Sept 2013.

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

[16] A. K. Sao, B. Yegnanarayana, and B. Vijaya Kumar. Significance of image representation for face verification. *Signal, Image and Video Processing*, 1:225–237, 2007.

[17] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1–8, June 2007.

[18] L. Sun and J. Hays. Super-resolution from internet-scale scene matching. In *IEEE International Conference on Computational Photography (ICCP), 2012*, pages 1–12, April 2012.

[19] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[20] C.-Y. Yang, J.-B. Huang, and M.-H. Yang. Exploiting self-similarities for single frame super-resolution. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision âĂŞ ACCV 2010*, volume 6494 of *Lecture Notes in Computer Science*, pages 497–510. Springer Berlin Heidelberg, 2011.

[21] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, Nov. 2010.

[22] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, volume 6920, pages 711–730. Springer, 2012.

[23] X. Zhang, E. Lam, E. Wu, and K. Wong. Application of Tikhonov regularization to super-resolution reconstruction of brain MRI images. In X. Gao, H. MÃijller, M. Loomes, R. Comley, and S. Luo, editors, *Medical Imaging and Informatics*, volume 4987 of *Lecture Notes in Computer Science*, pages 51–56. Springer Berlin Heidelberg, 2008.