Gradient Feature Based Image Super Resolution Using Neighbor Embedding

Pandya Hardeep[1], Prof. Prashant B. Swadas[2], Prof. Mahasweta Joshi[3]

[1]*Research Scholar Student, BVM, V V Nagar, GTU, India*
[2] *Head of Computer Department, BVM, V V Nagar, GTU, India*
[3] *Faculty in Computer Department, BVM, V V Nagar, GTU, India*

[1]hbpandya12@gmail.com; [2]pbswadas@bvmengineering.ac.in;[3]sweta.ce2013@gmail.com

***Abstract*-In this paper, we propose a novel method for solving single-image super-resolution problems which is also known as example based super resolution. Given a low-resolution image as input, we recover its high-resolution counterpart using a set of training examples. We have used manifold learning methods, and we have used first order and second order gradient based feature vector for each patch in the image. Locally linear embedding (LLE) method is used for manifold learning. Small image patches in the low- and high-resolution images form manifolds with similar local geometry in two discrete feature spaces. In LLE, local geometry is characterized by how a feature vector corresponding to a patch can be reconstructed by its neighbours in the feature space. Same way, in this method the training image pairs are used to estimate the high-resolution embedding. We have implemented local compatibility and smoothness constraints are used between patches in the target high-resolution image through overlapping. Experiments show that proposed method gives better results than previous methods with few training examples.**

***Keywords*—Single image super resolution, manifold learning, neighbor embedding, locally linear embedding, first and second order gradient.**

I. INTRODUCTION

A. Super Resolution

In digital image processing re-sizing of the image does not translate into an increase in resolution. We may call the process of re-sizing for the purpose of increasing the resolution as up sampling or image zooming, it is not super resolution technique but it is interpolation technique. Super-resolution (SR) refers to the task of producing a high-resolution (HR) image from one or more low-resolution (LR) images.

Super resolution techniques are mainly of two types. One is Multi frame image super resolution [1], [2] and another is Single frame image super resolution. In multi-frame super resolution, high resolution image is generated from registration of available low resolution images. In single-frame super resolution, high-resolution image is constructed from a single low-resolution image, with the help of a set of one or more training images from scenes of the same different types. It is also called as learning [8], [9], [10] or example based super resolution.

When there are not enough multiple images of same scene are available and magnification factor is high one cannot use multi-frame image super resolution, to overcome this problem learning based super resolution approach is useful. Learning based super-resolution, which recovers the high resolution (HR) by learning the relationship between HR training images and low resolution (LR) counterparts, attracts much attention recently [10], [11], [12], [13]. Comparing with other methods, learning based super-resolution is capable of extracting more image/patch information from a collection of image pairs or image patch pairs and supports higher magnification factors with fewer LR images.

Single-image Super-Resolution is often referred to as Example-based Super-Resolution as the prior information required in order to estimate the missing HR details is given in the form of examples, i.e. learned pairs of LR and HR patches (sub-windows of image) that compose a dictionary. The learned pairs of patches are usually taken from external HR images and degraded (blurred and conveniently downsized) versions of them. As a first step of any example-based SR algorithm, the target image is divided into patches of the same size of the LR patches in the dictionary; then, each LR input patch is compared to the stored LR patches and, once the nearest patch among these is found, the corresponding HR patch is finally taken as the output [10], [11], [12].

The image super-resolution problem arises in a number of real-world applications. Synthetic zooming of region of interest (ROI) is an important application in surveillance, forensic, scientific, medical, and satellite imaging. For surveillance or forensic purposes, a digital video recorder (DVR) is currently replacing the CCTV system, and it is often needed to magnify objects in the scene such as the face of a criminal or the licence plate of a car [1], [2].

Another application is found in web pages with images. To shorten the response time of browsing such web pages, images are often shown in low-resolution forms (as the so-called “thumbnail images”). An enlarged, higher resolution image is only shown if the user clicks on the corresponding thumbnail. In television also super resolution plays a major role in converting NTSC signal to a HDTV signal for high definition visual quality [3], [4].

B. Related Previous Work

Researchers have proposed many super resolution techniques which are either multi frame super resolution or single frame super resolution. The idea of super-resolution was first introduced in 1984 by Tsai and Huang [6] for multi-frame image restoration of band-limited signals. Most super-resolution methods are composed of two main steps: first all the images are aligned in the same coordinate system in the registration step, and then a high resolution image is reconstructed from the irregular set of samples. In this second step, the camera point spread function is often taken into account. Narasimha Kaulgud and U. B. Desai [5] have proposed image zooming using wavelets. Wavelets are most commonly used for multi resolution analysis. Some have proposed image super resolution technique in frequency domain with registration of aliased images. In that method they use registration algorithm for multiple images, from that they estimate rotation and shift estimation. Finally by removing aliasing they will get high resolution image. All these process is done in frequency domain [3] and for that they have used fourier transformation.

Simple resolution enhancement methods based on smoothing and interpolation techniques for noise reduction have been commonly used in image processing [1], [2], [5]. Smoothing is usually achieved by applying various spatial filters such as Gaussian, Wiener, and median filters. Commonly used interpolation methods include bicubic interpolation and cubic spline interpolation [13]. Interpolation methods usually give better performance than simple smoothing methods. However, both methods are based on generic smoothness priors and hence are indiscriminate since they smooth edges as well as regions with little variations, causing blurring problems. More recently, some learning-based methods have been proposed by different researchers. Some methods make use of a training set of images [13], [14], [15], [16]. The methods based on a training set are very similar in spirit. While the framework based on image analogies [10] was proposed for a wide variety of image texturing problems including super-resolution, the method is less effective than other super resolution methods as no interaction between adjacent elements (pixels or image patches) in the high resolution image is explicitly enforced to ensure compatibility. Nevertheless, all these methods use the training data in a similar way. In particular, each element in the target high-resolution image comes from only one nearest neighbor in the training set [15], [16].

A variation to this procedure is presented in [7] and in some other SR methods based on sparse representations [8] instead of selecting from the dictionary only one patch, several patches are taken into account and contribute simultaneously to the generation of a single HR output patch. In particular, in [11] the authors propose a single-image SR algorithm, based on the concept of neighbor embedding and originally inspired by a method for data dimensionality reduction called Locally Linear Embedding (LLE) [11]. The basic assumption is that a patch in the LR target image and the corresponding HR unknown patch share similar neighborhood structures: as a consequence of that, once the LR patch is expressed as the linear combination of a certain number of its neighbours taken from the dictionary, the output patch can be reconstructed by using the HR patches in the dictionary corresponding to the neighbors selected, and combining them in the same way. Also in some have proposed same method of [13] but they have used quaternion color image model. In that method neighbor embedding is applied to each of the RGB component value of a pixel and feature vector is also generated by first-order quaternion gradient. Recently a modified procedure to [10] is proposed in which they have considered same algorithm only for non negative matrix factor values. Means they have considered only non negative reconstruction weights only for the construction of high resolution image.

II. PROPOSED ALGORITHM

A. Neighbor embedding and LLE

The single-image super-resolution problem that we want to solve can be formulated as follows. First we are taking one high resolution image IH which we will use for training of input image and one low resolution image as test input image.

From available high resolution image we want to construct low resolution image for which we have applied filtering process to original high resolution image. After that we have sub-sampled output of that process to obtain the low resolution or degraded output image which is l_L. For generating same number of patches as high resolution image we have used bicubic interpolation to interpolate low resolution image which is I_{UP}. I_{UP} image has same resolution as I_H.

Any image has high frequency and low frequency component on it, the more high frequency part it has the better the quality of image. So, for obtaining high frequency and mid frequency component we have used band pass filtering approach which selects only limited band of frequencies. The HF image I_{HF} is obtained by subtracting I_{UP} from I_H. MF image is band pass version of I_{UP}.

For neighbor embedding method we will have used notation as input high resolution image is YS and XS is band pass version of I_{UP}. Next we have taken one low resolution test image as YT. We have obtained XT from YT by doing averaging operation on YT. Then XT is down sampled which is same as we performed with high resolution image.

The Neighbor Embedding (NE) algorithm [9], [11] is the starting point for this work. For neighbor embedding method several patches are taken from the dictionary to represent a single input patch; it follows that multiple patches, i.e. their HR counterparts in the dictionary, are combined together to generate an output HR patch. We will use feature vector for representing each patch. We have to extract one or more features from any single pixel of the patch, and then patch vector is then obtained by the simple concatenation of the features of its pixels. The nearest neighbor search, the reconstruction weight computation and the HR patch reconstruction steps are all performed with respect to the feature spaces selected. After the HR feature vectors are reconstructed, we finally "reverse" the features, so obtaining the actual HR output patches. In [11], for instance, a 4-value gradient and the centred luminance pixel values are used as features, respectively for the LR and HR patches. The following notations are used in Table 1 for corresponding feature vector representation. Here all patches are intended as patch feature vectors.

This method is based on the assumption that small patches in the low- and high-resolution images form manifolds with similar local geometry in two distinct spaces. This assumption is valid because the resulting representation is stable and hence independent of the resolution as long as the embedding is isometric. Each patch, represented as a feature vector, corresponds to a point in one of the two feature spaces. Locally Linear Embedding (LLE) is applied for manifold learning [11], [12].

The LLE algorithm is based on simple geometric intuitions. Suppose there are N points in a high-dimensional data space of dimensionality D, where the N points are assumed to lie on or near a nonlinear manifold of intrinsic dimensionality d < D. Provided that sufficient data points are sampled from the manifold, each data point and its neighbours are expected to lie on or close to a locally linear patch of the manifold. The local geometry of each patch can be characterized by the reconstruction weights with which a data point is reconstructed from its neighbours [11].

A feature vector corresponding to a patch can be reconstructed by its neighbors in the feature space. For each patch in the low-resolution image XT, first compute the reconstruction weights of its neighbors in Xs by minimizing the local reconstruction error. The high-resolution embedding is then estimated from the training image by preserving local geometry. Finally, local compatibility and smoothness constraints between adjacent patches in the target high-resolution image through overlapping are applied to obtain estimated high resolution image.

We are applying Locally Linear Embedding algorithm to each patch Xtq in image XT. First we have to find the set Nq of K nearest neighbours in Xs. Nq can be obtained by using Euclidean distance measure.

After obtaining Nq we have to compute reconstruction weights of the neighbours which minimize the error of reconstructing Xtq. This calculation provides optimal reconstruction weights as below.

$$\varepsilon^q = \| \mathbf{x}_t^q - \sum_{\mathbf{x}_s^p \in \mathcal{N}_q} w_{qp} \mathbf{x}_s^p \|^2$$

This is the squared distance between xtq and its reconstruction, subject to the constraints

So minimizing ε^q subject to the constraints,

$$\sum_{\mathbf{x}_s^p \in \mathcal{N}_q} w_{qp} = 1$$

$$w_{qp} = 0$$

$$\mathbf{x}_s^p \notin \mathcal{N}_q$$

It is a constrained least squares problem [11], [12]. For calculating reconstruction weights of Y_T we have to calculate Gram matrix for Xtq as below.

$$G_q = (x_t^q \mathbf{1}^T - X)^T (x_t^q \mathbf{1}^T - X)$$

where X is a D×K matrix with its columns being the neighbors of Xqt and 1 is a column vector of ones. By grouping the weights of the neighbours, it will form a K dimensional weight vector Wq by reordering the subscript p of each weight Wqp. The constrained least squares problem has the following solution.

$$w_q = \frac{G_q^{-1}\mathbf{1}}{\mathbf{1}^T G_q^{-1}\mathbf{1}}$$

Here calculation required for inverse of gram matrix is too much time consuming, to optimize it we have to calculate it in terms of linear system of equations.

$$\mathbf{G}_q \mathbf{w}_q = \mathbf{1}$$

After using above equation we have to normalize the weights which satisfies same conditions as above in least squares solution. We have to apply above procedures and equations to each patch Nq in X_T image. So finally we will get reconstruction weights which are obtained from weight matrix W.

$$W = [w_{qp}]$$

From this equation we now can calculate the initial value Ytq which is based on weight matrix W.

$$y_t^q = \sum_{x_s^p \in N_q} w_{qp} y_s^p$$

In above equation a procedure is applied to impose inter-patch relationships by averaging the feature values in overlapped regions between adjacent patches. One can implement other feature vectors also.

B. Feature Selection and Patch generation

In this method, each image patch is represented by a feature vector. Color images are commonly represented by the RGB channels. However, humans are more sensitive to changes in luminance than to changes in color. Therefore, instead of using the RGB color model, we use the YIQ color model where the Y channel represents luminance and the I and Q channels represent chromaticity. Conversion between the RGB and YIQ color schemes can be done easily via a linear transformation. In this method, the chromaticity components from the I and Q channels are not learned. They are simply copied from the low-resolution image to the target high-resolution image [11], [12]. Hence, only the luminance values from the Y channel are used to define features for each patch.

For the low-resolution images, one possible way is to define the feature vector as a concatenation of the luminance values of all pixels inside the corresponding patch. However, this simple method will not yield optimum performance. An alternative scheme, which we use here, is to consider the relative luminance changes within a patch [11], [13]. This feature representation scheme allows us to use a relatively small training set. More specifically, we have applied the first-order and second-order gradients of the luminance as features. Figure 1 shows a 5 × 5 local neighborhood of the pixel at the center with luminance value Z13. The first-order gradient vector of Z13, denoted ∇z13, and the second-order gradient vector, denoted ∇^2z13, can easily be derived as follows:

$$\nabla z_{13} = \begin{bmatrix} (z_{14} - z_{13}) + (z_{13} - z_{12}) \\ (z_{18} - z_{13}) + (z_{13} - z_{8}) \end{bmatrix}$$

$$= \begin{bmatrix} z_{14} - z_{12} \\ z_{18} - z_{8} \end{bmatrix}$$

$$\nabla^2 z_{13} = \begin{bmatrix} (z_{15} - z_{13}) - (z_{13} - z_{11}) \\ (z_{23} - z_{13}) - (z_{13} - z_{3}) \end{bmatrix}$$

$$= \begin{bmatrix} z_{15} - 2z_{13} + z_{11} \\ z_{23} - 2z_{13} + z_{3} \end{bmatrix}$$

z1	z2	z3	z4	z5
z6	z7	z8	z9	z10
z11	z12	**z13**	z14	z15
z16	z17	z18	z19	z20
z21	z22	z23	z24	z25

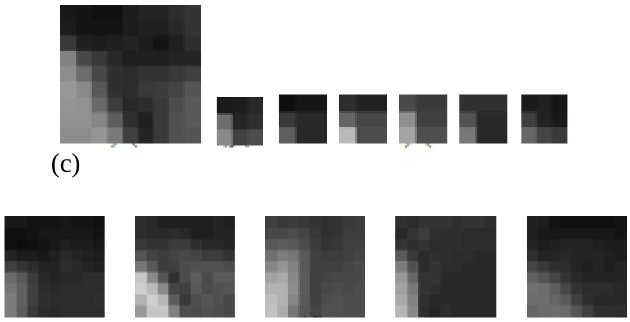
Figure 1: 5 × 5 local neighborhood of the center pixel is shown with luminance value z13.

In Chang et al. Method [11] they do not have considered diagonal pixel values while calculating first order and second order gradients which lacks of dimensionality. We have considered diagonal pixel values also in calculating first order and second order gradients of any pixel in 5 × 5 neighborhood.

So this equation provides more generalized way for calculating feature vectors of each patch in X_T^q. We have calculated above equations and we will get eight feature vectors corresponding to each pixel. The feature vector for each patch is then defined as the simple concatenation of the features for all pixels within the patch.

For the high-resolution images, we obtain the features for each patch based only on the luminance values of the pixels in the patch. Since the features used for the low-resolution patches cannot reflect the absolute luminance, we have to subtract the mean value from the luminance-based feature vector of each high-resolution patch. While we construct the target high-resolution image, the mean luminance value of the corresponding low-resolution patch will be added into it [11], [12].

Figure 2 shows an illustrative example of patch generation. The input low-resolution patch in (b) is down-sampled from a blurred version of the true high-resolution patch in (a). Using the feature selection described above, five nearest-neighbor patches in (c) are obtained from the training images and their reconstruction weights are computed according to previous equations. Based on the five corresponding high-resolution patches as shown in (d), the target high-resolution patch in (e) is constructed according to previous equations. The reconstructed high-resolution patch is perceptually very similar to the true high-resolution patch [11], [13]. None of the nearest-neighbor high-resolution patches is better than this reconstructed patch, showing the potential of this method in generalizing over the training images. This provides a sound justification for the satisfactory performance even when a small training set is used.



(a) (b) (c)

(d)

(e)

Figure 2: An illustrative example of patch generation

III. EXPERIMENTAL RESULTS

For generating high resolution image from low resolution image we have used following training images shown in figure 3 and figure 4 of size 280*280 in training dictionaries.

Figure 3: Training image

Figure 4: Training image

We have tested many low resolution test images and here we have shown some results of it. Here input image is (head.jpg) of size 140*140 as shown in figure 5 and result is shown in figure 6.

.

Figure 5: Input low resolution image

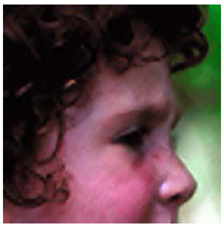


Figure 6: Output of proposed algorithm

As can be seen from figure constructed super resolution image gives better quality than just applying zooming to any low resolution image. Here we have shown result of only one image for our algorithm and shown PSNR calculations for various images.

We have calculated peak signal-to-noise ratio between input low resolution image and generated high resolution image. Results are shown in Table 1.

INPUT IMAGE	NEIGHBOR EMBEDDING	PROPOSED METHOD
Head.jpg	29.7845	30.3892
Tiger.jpg	32.4753	33.6689
Baboon.jpg	32.5321	33.1543
Lena.jpg	31.7423	32.3719
Model.jpg	29.1743	30.2281

Table 1: PSNR calculations

IV. CONCLUSION

As shown in table 1 proposed method yields better results than locally linear embedding algorithm. The higher the PSNR value the better the quality of image is obtained by this method. Here we have used training images in more general way. Here generation of a high-resolution image patch does not depend on only one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in a way similar to LLE for manifold learning. An important aspect is generalization over the training examples is possible and hence we can use fewer training examples in our method than other learning-based super-resolution methods. We have used first-order and second-order gradients of the luminance as features which can better preserve high-contrast intensity changes while trying to satisfy the smoothness constraints. Also we have considered diagonal pixels in calculations which have added more dimensionality and improved results. In future we can take multiple training images and can be applied to video shot super resolution.

References

[1] SubhasisChaudhuri (Indian Institute of Technology), " Super Resolution Imaging" Kluwer Academic Publishers, pp.1-44, 2002.

[2] Sung Cheol Park, Min KyuPark,and Moon Gi Kang, "Super-Resolution Image Reconstruction: A Technical Overview". IEEE Signal Processing Magazine May 2003.

[3] Patrick Vandewalle, Sabine Susstrunk and Martin Vetterli," A Frequency Domain approach to registration of aliased images with application to super-resolution" EURASIP journal on applied signal processing 2006.

[4] Gilman, A., Bailey, D.G., Marsland, S.R," Interpolation models for image super resolution", 4th IEEE International Symposium on Electronic Design, Test and Applications, DELTA 2008.

[5] NarasimhaKaulgud and Uday B. Desai, "Image Zooming: Use of Wavelets", The International Series in Engineering and Computer Science, Springer US Publishers,, ©Kluwer Academic Publishers.

[6] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. In R. Y. Tsai and T. S. Huang, editors, Advances in Computer Vision and Image Processing, volume 1, pages 317–339. JAI Press Inc., 1984.

[7] Jianchao Yang, John Wright, Thomas S. Huang, Yi Ma, "Image Super-Resolution as Sparse Representation of Raw Image Patches", IEEE conference on computer vision and pattern recognition, 2008.

[8] William T. Freeman, Thouis R. Jones, and Egon C. Pasztor, "Example-Based Super- Resolution" Image-Based Modeling, Rendering, and Lighting,IEEE March/April 2002.

[9] Mei GONG, Kun HE, Jiliu ZHOU, Jian ZHANG, "Single Color Image Super-resolution through Neighbor Embedding", Journal of computational information systems 7:1 (2011) 49-56.

[10] Marco Bevilacqua, AlineRoumy, Christine Guillemot, Marie-Line AlberiMorely, "Neighbor embedding based single-image super-resolution using semi-nonnegative matrix factorization", IEEE international conference on Acoustics, speech and signal processing, 2012.

[11] Hong Chang, Dit-Yan Yeung, Yimin Xiong, "Super-Resolution Through Neighbor Embedding", IEEE conference on computer vision and pattern recognition, 2004.

[12] Shin-Cheol Jeong and Byung Cheol Song, "Training based super resolution algorithm using k-means clustering and detail enhancement", published in 18th europien signal processing conference, august 2010.

[13] Kazuki Taniguchi, Motonori Ohashi, Xian-Hua Han, Yutaro Iwamoto, So Sasatani, Yen-Wei Chen, "Example based super resolution using locally linear embedding", Computer Sciences and Convergence Information Technology (ICCIT),6th International Conference on IEEE, 2011.

[14] Dattatray L. Karambelkar and P. J. Kulkarni, "Super resolution using manifold learning", published in Computational Intelligence and Communication Networks (CICN), International Conference on IEEE, 2011.

[15] Vishal R. Jaiswal, Girish P. Potdar, Tushar A. Rane,"Recent Developments in Super Resolution", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 2 No. 4, 2011.

[16] KathiravanSrinivasan, J. Kanakaraj, "A Study on Super-Resolution Image Reconstruction Techniques", Computer Engineering and Intelligent Systems, iiste, Vol 2, No.4, 2011