Advance Neighbor Embedding for Image Super Resolution

Dr. Ruikar Sachin D, Mr. Wadhavane Tushar D
Department of E&TC, Sinhgad Academy of Engineering, Pune, India
ruikaretcdept@gmail.com
tush.org@gmail.com

ABSTRACT

This paper presents the Advance Neighbor embedding (ANE) method for image super resolution. The assumption of the neighbor-embedding (NE) algorithm for single-image super-resolution Reconstruction is that the feature spaces are locally isometric of low-resolution and high-resolution Patches. But, this is not true for Super Resolution because of one to many mappings between Low Resolution and High Resolution patches. Advance NE method minimize the problem occurred in NE using combine learning technique used to train two projection matrices simultaneously and to map the original Low Resolution and High Resolution feature spaces onto a unified feature subspace. The Reconstruction weights of k-Nearest neighbour of Low Resolution image patches is found by performing operation on those Low Resolution patches in unified feature space. Combine learning use a coupled constraint by linking the LR–HR counterparts together with the k-nearest grouping patch pairs to handle a large number of samples. So, Advance neighbour embedding method gives better resolution than NE method.

Keywords

High Resolution (HR), Low Resolution (LR), grouping patch pairs (GPPs), combine learning, neighbor embedding (NE), super-resolution (SR).
INTRODUCTION

The Digital imaging system has lot of limitations, the imaging environment also important part of capturing an image, so it is not always easy to capture an image at a desired high-resolution (HR) level. However, in many practical applications such as medical imaging, video surveillance, computer vision, and entertainment, HR images are required usually to obtain a robust performance. Therefore, it becomes popular to synthesize a new HR image by using one or more low-resolution (LR) images [1].

The fast interpolation methods such as bilinear, bi-cubic, and other resampling methods [2]–[3] can increase the size of LR input, they are prone to blur high-frequency details and not super resolve the LR image, they only up-sample the image. Thus, a large number of super-resolution (SR) reconstruction techniques have been developed in recent years, which can be divided into two categories: multi image-based SR methods and example learning-based SR methods [4], [5].

There are two types of multi-image-based SR methods: The frequency-domain methods [6]–[7] and spatial-domain methods [8]–[9]. It has been proved that the spatial representation of super resolution gives better performance than the frequency domain one. Representative spatial-domain methods include the regularized SR reconstruction approaches [10]–[11], the projection-onto-convex-sets approach [12], iterative back projection (IBP) [13], adaptive filtering [14], nonlocal means [15]–[16], and nonlocal kernel regression [9]. But all the above mention methods require adequate number of LR images. Example learning based group of SR methods assumes that the high-frequency details lost in an LR image can be learned from a training set of LR and HR image pairs, i.e., the relationship between LR image patches and the corresponding HR patches can be used to estimate the missing HR frequency details in the given LR input. The example learning based method assumed that the counterparts of LR–HR image patch pairs (or their feature representations) are locally isometric. In contrast to above methods, this neighbor-embedding (NE)-based method does not require a large number of samples and achieves top level performance.

REVIEW OF NEIGHBOR EMBEDDING METHOD

Neighbor Embedding for SR reconstruction [19] is promising, except for its limitation of a locally isometric assumption in the Low Resolution and High Resolution feature spaces. To address this type of problem, the existing variations of the NE algorithm for image Super Resolution mainly concentrate on two aspects: one is to select more suitable features to characterize Low Resolution image patches such that the neighborhood relationship between LR–HR image patch pairs can be preserved as consistently as possible and the other is to build a better reconstruction function by imposing some consistency constraints on HR–LR pairs.

NE for SR reconstruction uses LLE [20] to estimate the HR patches corresponding to the LR inputs by assuming that the two manifolds in the HR and the LR image patch spaces are locally similar. Let $X_s = \{x_s^i\}_{i=1}^N$ be the training data set of the LR image patches, $Y_s = \{y_s^i\}_{i=1}^N$ be the training data set of the Corresponding HR image patch. And for Test image $X_t = \{x_t^i\}_{i=1}^M$ be the test LR Data set and $Y_t = \{y_t^i\}_{i=1}^M$ be the Estimated HR image patches. Here, $N$ and $M$ are the number of training image patches in the training data set and that of the test image patches in the test data set, respectively.

2.1 Algorithm 1 (NE for SR Reconstruction)

Input:
- Training data sets $X_s = \{x_s^i\}_{i=1}^N$ and $Y_s = \{y_s^i\}_{i=1}^N$
- Test data set $X_t = \{x_t^i\}_{i=1}^M$
- The number of nearest neighbor $k$

Output:
- $Y_t = \{y_t^i\}_{i=1}^M$

The neighbor embedding algorithm method can be summarized as follows:

1. For each patch $X_t^j$ in image $X_t$:
   a. Find the set $N_{x_t^j}(j)$ of $k$ nearest neighbors in $X_s$
   b. Compute the reconstruction weights of the neighbors that minimize the error of reconstructing $X_t^j$.
   c. Compute the high-resolution embedding using the appropriate high-resolution features of the $k$ nearest neighbors and the reconstruction weights.
2. Construct the target high-resolution image by enforcing local compatibility and smoothness constraints between adjacent patches obtained in step 1(c).
3. Step 1(a) implement by using Euclidean distance to define neighborhood. Based on the $k$-nearest neighbors identified, step 1(b) seeks to find the best reconstruction weights for each patch $x^j_t$ in $X_t$. Optimality is achieved by minimizing the local reconstruction error for $x^j_t$.

$$
\varepsilon^j = \min \| x^j_t - \sum_{i \in N_j(f)} w_{ij} x^i_t \|^2
$$

Which are the squared distance between $x^j_t$ and its reconstruction. Such that $\sum_{i \in N_j(f)} w_{ij} = 1$ and $w_{ij} = 0$ when $i \notin N_j(f)$.

ADVANCE NEIGHBOR EMBEDDING BASED SUPERRESOLUTION

In this section, a MAP reconstruction framework for SR is discussed. A coupled constraint on $k$-NNs of LR–HR counterparts, i.e., GPPs (groping patch pairs), is then established for the combine learning process. Thereafter, for each LR image patch to be super-resolved, its nearest GPPs (including $k$-NNs of LR–HR image patch pairs) is searched to perform combine learning for the unified feature subspace. Subsequently, the selection of $k$-NNs and the optimal weights for reconstruction are performed in the unified feature subspace for the initial HR output patches. Fig. 1 illustrates the reconstruction framework of Advance NE algorithm.

- The coupled constraint on GPPs consisting of the LR–HR image patches is applied to the learning process, which ensures a better consistency local LR and HR image patches.
- Matrices $P_l$ and $P_h$ are the mapping matrices obtain from combine learning such that the difference between LR–HR counterparts is reduced as much as possible in the unified feature subspace. The optimal reconstruction weights for SR reconstruction are then estimated in the unified feature subspace rather than solely in the LR feature space.

3.1 MAP Reconstruction Framework

HR images can be divided into three sub-bands, the high frequency band $I_h$, the middle frequency band $I_m$ and the low frequency band $I_l$. The middle-frequency components can be recovered based upon a simple interpolation algorithm (e.g., bilinear or bicubic). Thus, the goal of SR reconstruction is to estimate the missing high-frequency component with a single LR input. If prior knowledge, which is denoted as manifold MF, can be incorporated into the SR process, the problem of SR reconstruction can be written as,

$$
I^*_h = \arg \max_{I_h} P(I_h, I_I, MF)
$$

By using Beyes rule, (1) becomes,

$$
I^*_h = \arg \max_{I_h} P(MF / I_h, I_I)P(I_I / I_h)P(I_h)
$$

Taking the negative log likelihood $E = -\log (p (.))$ in (2), we can get,

$$
I^*_h = \arg \min_{I_h} E(MF / I_h, I_I) + E(I_I / I_h) + E(I_h)
$$
Where $f_j$ and $f_h$ denote two mapping functions used to convert two different feature vectors $x^j_i$ and $y^j_i$ into a unified feature space. In this way, use a common distance metric to measure the similarity between them.

The term $E(I_f / I_h)$ stands for the data cost of the reconstruction error of all input LR image patches with respect to their neighbourhood $N_g(.)$ i.e.

$$E(I_f / I_h) = \sum_{j} \left\| f_j(x^j_i) - \sum_{i \in N_g(j)} w_{ij} f_h(y^j_i) \right\|^2$$ (4)

The term $E(I_h)$ can be seen as the smoothness prior of the HR image, which can be approximated by overlapping one or several pixels within the adjacent image patches. Therefore, the key problem to solve is finding the two mapping functions $f_j$ and $f_h$ to transform the LR and HR feature spaces into a unified feature subspace. Once achieved, utilize the traditional NE algorithm to estimate the optimal weights and generate the desired HR image patch as a linear combination of the corresponding HR image patches from its neighbors. In addition, the problem of finding $M_j$ associated with each $x^j_i$ remains unresolved. This problem is addressed by grouping the $k$-nearest group patch pairs of the LR and HR features in the training data set.

### 3.2 Coupled Constraint

In NE method [17], the training data set of the LR image patches, $X = [x^i_s]_{i=1}^N$ and that of their corresponding HR image patches $Y = [y^i_s]_{i=1}^N$ where represents the index of each image patch and $N$ is the number of image patches collected. In the traditional NE for SR $x^i_s$ represents a $d$-dimensional feature vector of the $i$th LR image patch by concatenating the first- and second-order gradient features in horizontal and vertical directions, respectively. The $y^i_s$ is the high frequency details of the $i$th HR image patch. In order to apply combine learning to two different feature spaces of LR and HR patches, and augment another training data set $Z = [z^i_s]_{i=1}^M$ that consists of the first- and second-order gradient features of HR image patches in horizontal and vertical directions, which is similar to the representation of each LR image patch. In such way $z^i_s$, denotes an $m$-dimensional feature vector of the $i$th HR image patch.

Most existing NE algorithms perform SR reconstruction without considering the correlation between the LR and HR image patches. That means that they solely apply an isometric assumption to synthesize the expected HR estimates. However, this assumption does not hold perfectly for the SR problem. To solve or at least to reduce this problem, Advance NE considers the similarly between the LR image patch and the HR image patch with a coupled constraint as follows.

Let $C = [c_i]_{i=1}^N$ be a coupled set by concatenating each feature vector $x^i_s$ and $y^i_s$ . Thus, each column measurement $c^i$ in set $C$ is a ($d+m$) dimensional feature vector, i.e.

$$c^i = \left[ \frac{x^i_s}{\sqrt{d}} \right] = \left[ \frac{y^i_s}{\sqrt{m}} \right]$$ (5)

To adjust the dynamic changes of the concatenated feature Vector, normalize it to a unit two-norm. For each vector $C$ in Set C, select the k-NNs associated with it and group them Together by

$$G^i = \{C_j\}_{j \in N_k(i)}$$ (6)

Where $G^i (1 \leq i \leq N)$ stands for the i-th GPPs related to $x^i_s$ and $N_k(i)$ represents the index set of k-NNs of $c^i$.
3.3 Combine Learning on Patch Pairs

Under the MAP reconstruction framework described in Section II-A, find the two mapping functions \( f_l \) and \( f_h \) to project the LR and HR feature spaces onto a unified feature subspace. Then apply combine learning to transform two feature spaces spanned by different dimensional feature vectors into a unified feature subspace.

Let \( L^i = R^{d_i} (i = 1, ..., N) \) denote the constraint patch matrix by stacking \( K \) \( d \)-dimensional column vectors whose indices are specified in \( G_i \), i.e., \( L^i = \{x_s^i\}_{s \in G_i} \). Similarly, \( H^i = R^{m_i} \) and \( H^i = \{z_s^i\}_{s \in G_i} \). The existing NE-based SR [17] measures the similarity in the LR feature space. Generally, these algorithms make an assumption that the relationship between LR and HR feature spaces is locally isometric, i.e., the neighbor relationship between the LR and HR feature spaces is locally preserved.

LR and HR features can share a unified feature subspace in which they are more closely associated with each other as in [18]. Based on this consideration, the feature representations of LR and HR counterparts are projected onto a unified feature subspace by combine learning via a coupled constraint. Following this, the selection of k-NNs is conducted in this unified feature subspace instead of the original LR space. The combine learning on the GPP of each LR input locally also performed, rather than on all training samples, which is efficient and tractable for a training data set containing an enormous number of samples.

In unified feature space two mapping matrices are measured one is for the LR feature space \( f_l = R^D \rightarrow R^P \) and another is for HR feature space \( f_h = R^m \rightarrow R^P \). In this way, project the LR and HR feature spaces simultaneously onto a unified feature subspace and measure their similarity by

\[
d_{ij} = \text{Dist}(f_l(x_s^i), f_h(z_s^i))
\]

(7)

Whereas function distance (\( \text{Dist}(.) \)) represent Euclidean distance.

For (7), the problem is converted to construct the mapping functions \( f_l \) and \( f_h \) such that distance \( d_{ij} \) should be as close as possible in the unified feature subspace. If the similarity is measured by the Euclidean distance, then

\[
\arg \min_{\{f_l, f_h\}} \sum_{i \in G} \left\| f_l(x_s^i) - f_h(z_s^i) \right\|_2^2
\]

(8)

Suppose the two mapping functions \( f_l \) and \( f_h \) are represented by projection matrices \( P_l \in R^{d \times p} \) and \( P_h \in R^{m \times p} \) then eq. (8) can modify as-

\[
\arg \min_{\{P_l, P_h\}} \sum_{i \in G} \left\| P_l^T x_s^i - P_h^T z_s^i \right\|_2^2
\]

\[
\arg \min_{\{P_l, P_h\}} \text{tr} \left( P_l^T L' 0 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} P_h^T \begin{bmatrix} L' & 0 \\ 0 & H' \end{bmatrix} \begin{bmatrix} P_l^T P_l & \end{bmatrix} \right)
\]

(10)
\[
\arg\min_{\mathbf{P}} \text{tr}\left( \mathbf{P}^T \mathbf{S} \mathbf{A} \mathbf{S}^T \mathbf{P} \right) \tag{11}
\]

This equation (11) obtain by assigning P, S, A to matrices shown in eq. (10) respectively.

Suppose that \( \mathbf{E} = \mathbf{S} \mathbf{A} \mathbf{S}^T \) and \( \mathbf{F} = \mathbf{S} \mathbf{S}^T \), the optimization problem with respect to P can be obtained by the eigenvectors \( \mathbf{p} \) of \( \mathbf{E}_p = \lambda \mathbf{F}_p \) associated with the second to the \( p \)-th \( (P \leq d) \) smallest eigenvalues. Here, matrices E and F are of size \((d + m) \times (d + m)\). For the solutions to \( p_l \) and \( p_h \), expand the \( \mathbf{E}_p = \lambda \mathbf{F}_p \) to two linear equations, i.e.,

\[
\mathbf{L}'(\mathbf{L})^T \mathbf{p}_l - \mathbf{L}'(\mathbf{H})^T \mathbf{p}_h = \lambda \mathbf{L}'(\mathbf{L})^T \mathbf{p}_l \tag{12}
\]

\[
\mathbf{H}'(\mathbf{H})^T \mathbf{p}_h - \mathbf{H}'(\mathbf{L})^T \mathbf{p}_l = \lambda \mathbf{H}'(\mathbf{H})^T \mathbf{p}_h \tag{13}
\]

From (13), the solution to \( p_h \) can be obtained from,

\[
\mathbf{p}_h = \frac{(\mathbf{H}'(\mathbf{H})^T)^{-1} \mathbf{H}'(\mathbf{L})^T \mathbf{p}_l}{(1 - \lambda)} \tag{14}
\]

According to (14), eliminate \( p_h \) in (12) and obtain

\[
\mathbf{H}'(\mathbf{H})^T (\mathbf{H}'(\mathbf{H})^T)^{-1} \mathbf{H}'(\mathbf{L})^T \mathbf{p}_l = (1 - \lambda)^2 \mathbf{L}'(\mathbf{L})^T \mathbf{p}_l \tag{15}
\]

Let eq. (15) reduces to (16) by replacing, U and V.

\[
| U_p | = (1 - \lambda)^2 | V_p | \tag{16}
\]

Once \( \mathbf{p}_l \) is obtained, substitute it into (13) for \( \mathbf{p}_h \). By selecting the appropriate dimension of the \( p \) subspace (or the unified feature subspace), use the two derived projection matrices \( \mathbf{p}_l \) and \( \mathbf{p}_h \) to transform the original LR and HR feature spaces into a common subspace. In this way, the selection of k-NNs can be conducted within the unified feature subspace. Combine learning for the projection matrices is summarized in algorithm.

### 3.4 Algorithm 2 (Combine Learning for Projection Matrices)

**Input parameters:**
- Training data set \( X_s = \{x'_i\}_{i=1}^{N} \) and \( Z_s = \{z'_i\}_{i=1}^{N} \)
- GPPs set \( \{G'_i\}_{i=1}^{N} \);
- LR patch \( x'_i \) in the test data set \( X_t \)

**Output parameters:**
- Projection matrices \( \mathbf{P}_l \) and \( \mathbf{P}_h \);
- Constraint patches matrices \( \mathbf{L}' \) and \( \mathbf{H}' \)

**Description:**
1. Find the nearest neighbour \( x'_i \) related to \( x'_j \) in the training data set \( X_s \).
2. Obtain the GPPs \( G'_i \) corresponding to \( x'_j \).
3. According to the index set specified in \( G'_i \), build the constraint patch matrices \( \mathbf{L}' \) and \( \mathbf{H}' \) respectively.
4. Construct projection matrices \( \mathbf{P}_l \) and \( \mathbf{P}_h \).

### 3.5 Algorithm 3 (Advance Neighbor Embedding)

**Input parameters:**
- Training data set \( X_s = \{x'_i\}_{i=1}^{N}, Y_s = \{y'_i\}_{i=1}^{N} \) and \( Z_s = \{z'_i\}_{i=1}^{N} \)
- GPPs set \( \{G'_i\}_{i=1}^{N} \); LR image ;
- The size of LR image patch \( q^*q \);
- Neighborhood size \( k \);
- Dimension \( P \) of the unified feature subspace to be projected.
Output parameters:
- HR image Y

Description:
1. Partition Y into q*q image patches with one or two pixels overlapped in raster-scan order to construct the test data set \( X_t = \{ x_i^j \} \) 

2. For each patch \( x_i^j \) in \( X_t \), execute the following steps repeatedly:
   - Compute the mean values \( \bar{x} \) of patch \( x_i^j \)
   - Construct the constraint patch matrices \( P_i \) and \( P_h \) related to the nearest neighbour \( x_i^j \) and compute the projection matrices \( P_i \) and \( P_h \) associated with \( x_i^j \) with algorithm 1.
   - Compute the transformed feature of \( x_i^j \) via \( P_i^T L \) and \( P_h^T H \) respectively.
   - Compute the optimal weights by minimizing reconstruction error as follows:
     \[
     \arg \min_{\{w_{ij}\}} \sum_{i \in N_x^j} \left( P_i^T x_i^j - \sum_{i \in N_y^j} w_{ij} P_h^T H \right)^2 \quad (17)
     \]
   - Reconstruct each \( y_i^j \) corresponding to \( x_i^j \) with the optimal weights \( w_{ij} \) as follows:
     \[
     y_i^j = \sum_{i \in N_y^j} w_{ij} y_i^j \quad (18)
     \]
   - Sum up mean values \( \bar{x} \) and \( \bar{y}_i^j \) together to generate the HR patch \( y_i^j \) and append it to \( Y_t \).

3. Produce the initial HR image \( X_0 \) by merging all the HR patches in the set \( Y_t \) for the overlapping region between those adjacent patches, averaging fusion is applied to obtain the estimated pixels.

**EXPERIMENTAL RESULT**

The PSNR obtained for the different images is given in table below. And the SR results of different images shown in fig. 3, 4, and 5 respectively.

| Table 1. PSNR |
|-------|---------------|----------------|
|       | Bi-Cubic Interpolation (db) | Neighbor Embedding (db) | Advance NE (db) |
| Boy.bmp (255*258) | 32.699140 | 39.7584 | 33.279377 |
| Grapes.bmp (306*306) | 34.154598 | 41.534 | 35.334959 |
| Fruit.bmp (360*360) | 31.328248 | 37.887 | 32.821744 |

Fig. 3 shows results for Boy.bmp (255*258) image. Upsample (b), (c), (d) by factor *3.

Fig.3 (a) is down-sampled by *3 version of original image. Fig.3 (b) is High-Resolution image obtained from Bicubic Interpolation method. Fig.3 (c) is High-Resolution image obtained from Neighbor Embedding method. Fig.3 (d) is High-Resolution image obtained from Advance Neighbor Embedding Method and Fig.3 (e) is original High-Resolution image.
Fig. 4 shows results for Grapes.bmp (306*306) image. Upsample (b), (c), (d) by factor *3. 

Fig.4 (a) is down-sampled by *3 version of original image. Fig.4 (b) is High-Resolution image obtained from Bicubic Interpolation method. Fig.4 (c) is High-Resolution image obtained from Neighbor Embedding method. Fig.4 (d) is High-Resolution image obtained from Advance Neighbor Embedding Method and Fig.4 (e) is original High-Resolution image.

CONCLUSION

This paper represents Advance Neighbor Embedding and conventional NE-based single-image SR reconstruction. Thorough experimental results demonstrate that Advance NE method gives the best SR output. The distinction of the advance NE lies in the selection of k-NNs and the linear embedding being performed in the unified feature subspace rather than in the original LR space. In order to obtain the formation process of the LR image from its HR version, all the training HR images are downsampled by using the bicubic interpolation by a factor of 3 to obtain the corresponding LR images. Since the human visual system (HVS) is more sensitive to the luminance component than the chrominance components. Use of the YCbCr color space for color images and only perform SR reconstruction in the luminance component. Considering that the middle-frequency information of LR images has greater correlation with high frequency than low frequency, first magnify the original LR input by a factor of 3 with the bicubic interpolation and then perform SR on it. In this implementation, the combine learning technique on GPPs those are most relevant to each LR input patch to obtain the desired unified feature subspace. In principle, the Advance NE method has the potential to be extended to other SR applications such as face image hallucination. In addition, the construction of optimal GPPs rather than a fixed neighborhood size is challenging for this method.

When the magnification increase (e.g., more than *3 up-scaling), the SR problem becomes severely undetermined. It is also important to notice that NE methods takes bit more time to execute. Only single training image is used in NE to recover patches, but in Advance NE training dictionary means multiple training images used to recover the patches to obtain HR ouput.

REFERENCES

Ruikar Sachin D had received Graduate degree in Electronics and Telecommunication from Government Engineering College, Aurangabad, under the aegis of Dr. B. A. M. U., Aurangabad, in 1998. He had received postgraduate degree in Electronics and Telecommunication Engineering from Government Engineering College, Pune University, India in 2002. He had completed PhD in Electronics under Shri Guru Gobind Singh Institute of Engineering Technology, SRTMU Nanded, in 2013. Presently, he is working as Assistant Professor in Electronics and Telecommunication at Sinhgad Academy of Engineering, Pune. His research interests include image denoising with wavelet transforms, image fusion, image painting and Image Super resolution.

Tushar Wadhavane had received Graduate degree in Electronics and Telecommunication from Sanjivani Rural Education Society’s College of Engineering, Kopargaon, Dist. A.Nagar, India, under aegis of Pune University in 2011. He has pursuing postgraduate degree in Electronics Engineering from Sinhgad Academy of Engineering, University of Pune