

# An Example Based Non-local Mean Regularization For Image Super-resolution

Chuanhai Li<sup>1</sup>, Ruimin Hu and Yang Xia

**Abstract.** Image super-resolution (SR) is a very useful technique for visual surveillance, high-definition TV and medical image processing. In the traditional method, the inferred high-resolution image patch is represented as a linear combination of dictionaries obtained by training images, and least square estimation or sparse regularization is used to find better solution. However, the methods proposed so far neglect the prior knowledge about local structure similarity in natural images. In this paper we introduce an adaptive regularization terms into LLE based SR framework. First, an example based image nonlocal mean regularization term is learned from the dataset of example image patches, which captures the local structure similarity between the input image and training images. Then, the image nonlocal self-similarity is used as another regularization term. In addition, we propose an iterative optimization framework to find the latent HR image. Experimental results on real surveillance images demonstrate the superiority of the proposed method over some state-of-the-art image SR approaches.

**Keywords:** Image super-resolution, nonlocal mean regularization..

---

<sup>1</sup> Chuanhai Li (✉)  
National Multimedia Software Engineering Research Center, Wuhan University, Wuhan, China  
China Telecom Wuhan Branch, Wuhan, China  
e-mail: 231760358@qq.com

Ruimin Hu  
National Multimedia Software Engineering Research Center, Wuhan University, Wuhan, China

Yang Xia  
National Multimedia Software Engineering Research Center, Wuhan University, Wuhan, China

## 1. Introduction

Image super-resolution (SR), referring to the technique of generating a high-resolution (HR) image from a low-resolution (LR) one, is very useful for visual surveillance, high-definition TV and medical image processing. Because learning based SR approaches can take advantages of additional and relevant information from training images to achieve higher performance, there have been a lot of promising and encouraging developments in this field recently.

The common idea of learning based SR approaches is to infer HR image by learning the relationship between HR training images and corresponding LR ones. Baker et al. [1] used a Bayesian framework to infer the high-frequency components from a parent structure, which captured the relationship between the images of different scales. Wang et al. [2] proposed a learning-based SR method based on principal component analysis (PCA), which fitted the input face image as a linear combination of LR training face images so that the combination weights were used to reconstruct HR face image. However, since facial image contains abundant information and each has its own character, the global reconstruction based on PCA tends to miss some high-frequency details, especially around the contours of the face and mouth. Therefore, some researchers explored local reconstruction methods based on small image patches [3-5]. Inspired by Locally Linear Embedding (LLE), Chang et al. [3] assumed that small image patches in LR and HR images form manifolds with similar local geometry in two distinct feature spaces. So, the HR image patch was reconstructed by its neighbors in the feature space, using the local geometry learned via input LR image patch and its neighbor patches. However, due to a mass of example patches used to represent the feature space, the complexity of Chang's method is very high. When reducing the number of example patches for some low-complexity applications, such as face SR on mobile platform, the representation ability of the example patches also decreases, and some artifacts may appear in the inferred HR face image. Ma et al. [4-5] utilized position-patches to reduce the search area of the latent neighbor patches, which were defined as the patches in all training images that have the same position with certain input image patch. For position-patches are usually with a structure more similar to input image patch, the approach is more efficient and excellent compared with Chang's method [3]. However, the dimension of image patch is usually smaller than the number of the training pictures, which may lead to overfitting of linear combination.

To address the problem of overfitting or poor fitting, prior knowledge of natural images has been used to regularize SR problem. The total variation (TV) model [6] is the most commonly regularization tool, but it tends to smooth out the fine details of an image. Yang et al. [7] used compressed sensing based regularization term to solve SR problem, which replaced the least square estimation used in the work [3] with a convex constrained optimization to obtain more well-pleasing result. However, it neglects the prior knowledge about local structure similarity of

natural images, which was used to alleviate the effect of motion estimation error in video super-resolution and achieved more robust performance [8].

In this paper, we not only obtain the prior knowledge from the input image, but also explore it from example images. We introduce an example based image nonlocal mean (NLM) regularization into LLE based SR framework, which captures the local structure similarity between the input image and training images. At meanwhile, an image nonlocal self-similarity is also used as another regularization term. To solve the SR problem we propose an iterative optimization framework to find the latent HR image. The rest of the paper is organized as follows. Sec. 2 presents the proposed SR algorithm. Sec. 3 shows experiment results of proposed and some referenced algorithms on real surveillance images. Sec. 4 concludes the paper.

## 2 The proposed SR Approach

In this paper, we use NLM to capture local structure similarity between the input image and training images. At first, the example based non-local mean penalty function is introduced.

### *2.1 Example Based Non-local Mean penalty function*

The NLM has been posed originally in [9] for image denoising, in which every pixel is replaced with a weighted average of its neighborhood. The NLM weights are evaluated using block-matching fit between image patches centered around the center pixel to be filtered, and the neighbor pixels to be averaged. Suppose the latent HR image  $Z$  is given.  $Zp_i$  denotes a pixel at location  $i$ , and the filtered HR image pixel  $Zp_i^*$  can be obtained by:

$$Zp_i^* = \frac{\sum_{j=1}^J Zp_j * nw_{i,j}}{\sum_{j=1}^J nw_{i,j}} \quad (1)$$

where  $nw_{i,j}$  is the NLM weight between pixel at location  $i$  and its neighbor at location  $j$ , and  $J$  is the number of neighbor pixels. It can be observed that  $Zp_i^*$  depends on the value of the patch center pixel and the values of the neighbor pixels around the center pixel, which can capture the local structure similarity.

Considering there is much nonlocal redundancy in natural images, we expect the following prediction error  $E_{nlm}(Z)$  should be small, which can be used as regularization term.

$$E_{nlm}(Z) = \sum_{i=1}^V \left\| Zp_i - \sum_{j=1}^J Zp_j * nw_{i,j} / \sum_{j=1}^J nw_{i,j} \right\|_2^2 \quad (2)$$

where  $V$  is the number of pixels in the HR face image.

In this paper, we also use the local structure similarity between the input image and training images, which form the regularization term similar to formula (2).

$$E_{enlm}(Z, YH^n) = \sum_{i=1}^V \left\| Zp_i - \sum_{k=1}^K ew_{i,k} YHp_k / \sum_{k=1}^K ew_{i,k} \right\|_2^2 \quad (3)$$

where  $E_{enlm}(Z, YH^n)$  is prediction error between the HR image and the estimated image averaged by example images,  $YHp_k$  is a pixel at index  $k$  of  $Zp_i$ 's neighbor pixels, which are searched from example images  $YH^n$  ( $n=1, \dots, N$ , where  $N$  is the training image number), and  $ew_{i,k}$  is the NLM weight between  $Zp_i$  and its neighbor pixel with index  $k$ ,  $K$  is the number of neighbor pixels in the training images.

## 2.2 NLM weights Caculation

We expect the NLM weights should depend on the Euclidean distance between the true patches [10]. Let  $XH$  denotes the original HR image,  $XHnb_i$  is the patch centered around the center pixel at location  $i$ , and  $YHnb_k$  is the patch centered around the neighbor pixel  $YHp_k$  ( $k$  is the pixel index). The Euclidean distance  $\|XHnb_i - YHnb_k\|_2^2$  should be used to calculate  $ew_{i,k}$  and the Euclidean distance  $\|XHnb_i - XHnb_j\|_2^2$  should be utilized for  $nw_{i,j}$ . Therefore, the Euclidean distance between the true patches can be calculated by (4) and (5).

$$\|XHnb_i - YHnb_k\|_2^2 = E \left( \|Znb_i - YHnb_k\|_2^2 - \sigma^2 * S^2 \right) \quad (4)$$

$$\|XHnb_i - XHnb_j\|_2^2 = E \left( \|Znb_i - Znb_j\|_2^2 - 2\sigma^2 * S^2 \right) \text{ s.t. } i \neq j \quad (5)$$

where  $Znb_i$  is patch centered around  $Zp_i$ ,  $\sigma^2$  is variance of the noise, and  $S$  is the patch size of  $Znb_i$ , which is selected to 15 in our paper.  $E(\cdot)$  refers to mathematical expectation.

Based on the equality (4), the weights  $ew_{i,k}$  can be defined as:

$$ew_{i,k} = \exp \left( -1 / (hp * S^2 * (\|Znb_i - YHnb_k\|_2^2 - \sigma^2 * S^2)) \right) \quad (6)$$

where  $hp$  is a controlling factor and is 100 in this paper. And the self-similarity

weights  $nw_{i,j}$  can be obtained by

$$nw_{i,j,i \neq j} = \exp\left(-1 / (\hbar p * S^2 * \left(\|Znb_i - Znb_j\|_2^2 - 2\sigma^2 * S^2\right))\right) \quad (7)$$

when  $i=j$ , the self-similarity weights  $nw_{i,j}$  equals to 1.

### 2.3 Image SR based on NLM Regularization

Let  $X, Z, YL^n, YH^n$  denote the input LR image, the estimated HR image, LR training face image and HR training face image respectively.  $X, Z, YL^n, YH^n$  can be represented as a set of small overlapped image patches, which are denoted as  $Xsb_m, Zsb_m, YLsb_r, YHsb_r$ . The variable  $m=1, \dots, M$ , where  $M$  is the patch number. The variable  $r=1, \dots, R$ , where  $R$  is the sample number.

Because of the structural similarity, the input image patches can be reconstructed from the optimal linear combination of the training images patches. In [3], each patch  $Xsb_m$  can be represented by:

$$Xsb_m \cong \sum_{r=1}^L w_r * YLsb_r \quad (8)$$

where  $w_r$  represents the contribution of each patch  $YLsb_r$  to the reconstruction of the input face image's patch  $Xsb_m$ . And

$L$  is the number of neighbor pixels.

In this paper, we introduce two adaptive regularization terms to the formula (9).

$$Z^*, w_m^{n*} = \arg \min_{w_m^n, Z} \left\| Zsb_m - \sum_{n=1}^N w_m^n * YHsb_m^n \right\|_2^2 + \lambda_1 E_{enlm}(Z, YH^n) + \lambda_2 E_{nlm}(Z) \quad (9)$$

where  $\lambda_1$  and  $\lambda_2$  are constant balancing the contribution of nonlocal regularization, and  $\lambda_1 = \lambda_2 = 1$  in this paper. Eq. (9) can be solved by an alternative iterative optimization to find optimal represent coefficients and optimal HR image.

We use an iterative optimization framework to find the latent HR image, and the proposed image SR method is summarized as follows:

---

#### Algorithm 1: The Proposed SR Scheme

---

Input: LR face image, LR training images, HR training images

1. Calculate the initial HR face image  $Z_{initial}$  by following formula:

$$Z_{initial} \cong \sum_{r=1}^L w_r * YHsb_r$$

2. For  $u = 1, \dots, U$ ,  $U$  is the number of iteration
-

- 
- For  $i=1, \dots, V$ ,  $V$  is the number of pixels of  $Z$
3. Find the neighbor patches in  $Z_u$  and  $YHbb_{i,u}^q$
  4. Calculate  $ew_{i,k,u}^{pixel}$  and  $nw_{i,i,u}^{pixel}$  using (6) and (7)
  5. Update  $Z$  by following formula
 
$$Zp_{i,u+1} = Zp_{i,u} - \tau^* \begin{pmatrix} Zp_{i,u} - Z_i^{Initial} + \lambda_1 \left( Zp_{i,u} - \sum_{k=1}^K ew_{i,k,u} YHp_{k,u} / \sum_{k=1}^K ew_{i,k,u} \right) \\ + \lambda_2 \left( Zp_{i,u} - \sum_{j=1}^J nw_{i,j,u} Zp_{j,u} / \sum_{j=1}^J nw_{i,j,u} \right) \end{pmatrix}$$

where  $\tau = 0.001$  in this paper.

end
  6. update the search window of  $Znb_i$  in  $YH_{i,q}^{sp}$  (in practice, we update the search window of  $Znb_i$  when  $\text{mod}(u, H) = 0$ ,  $H = 100$ )
 

end
- Output: the estimated HR image  $Z_u$
- 

### 3 Experiment and Result

In the experiments, we perform the proposed method on real surveillance images, which is taken with a CIF-size ( $352 \times 288$  pixels). As shown as Fig.1 (left), the low-quality picture is taken by a surveillance camera in a dark room and the interested object is at a distance. We also capture the high-quality picture as shown as Fig.1 (right), which is taken in the normal light and at a short distance, to verify the effectiveness of the proposed method.



Figure 1. Example of the low-quality picture and high quality picture taken by surveillance camera

In the proposed method, the frontal and pre-aligned images of CAS-PEALR1 database [11] are used. There are 30867 images of the 1040 subjects in CASPEAL-R1 database. We carry out experiments on 500 images with neutral expression and

frontal pose, which are selected randomly for training. The images are aligned and cropped to  $112 \times 96$  pixels. The LR images are formed by smoothing (an averaging filter of size  $4 \times 4$ ) and down-sampled (by a factor of 4, thus the size of LR face images are  $28 \times 24$  pixels).

We compare the proposed method with Chang's LLE based SR method [3], and Yang's SR method [7]. For the sake of fair competition, we set the HR patch size to  $16 \times 16$  in all referenced methods and our proposal, and the overlap between

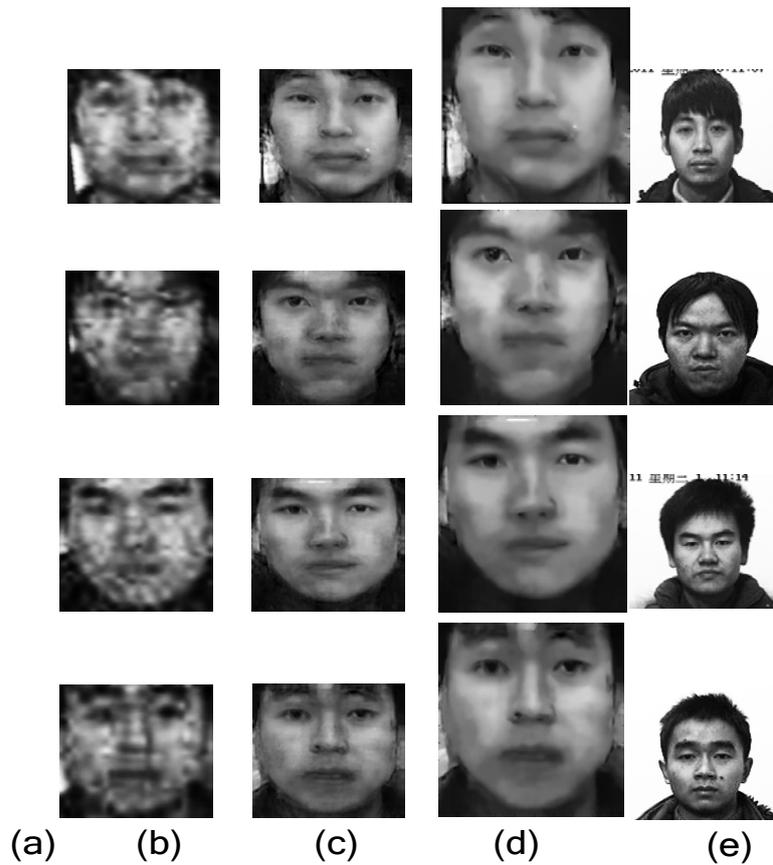


Figure 2. Comparison of results. (a) Input LR faces. (b) Chang et al. [3] (c) Yang et al. [7] (d) Our method. (e) Original HR faces.

The experimental results of test image are shown in Fig.2. Due to space limitation, we don't provide all the results here. From Fig.2 (b), we can see that Yang's method cannot distinguish the true image and noise, the synthesized HR face image contains some obvious noise and the hallucinated face images by Chang's method have some artifacts, primarily due to under-fitting. A further examination shows that our method (e) successfully exploit more information about local structure from the natural face images, and the inferred HR images appear smoother than

Chang's and Yang's ones. The excellent visual quality of our method owes to using the image local structure similarity contained in the natural images.

## 4 conclusion

In this paper we introduce an adaptive regularization terms into LLE based image framework to explore the prior knowledge about local structure similarity. An example based image nonlocal mean regularization term that captures the local structure similarity between the input image and training images and an image non-local self-similarity regularization term are used to improve the quality of estimated HR image through iterative optimization. Experimental results demonstrate the superiority of the proposed method over some state-of-the-art image SR approaches, especially for noise images.

**Acknowledgement:** This work is supported by the major national science and technology special projects (2010ZX03004-003-03); the National Natural Science Foundation of China under Grant No. 61172173, 60970160, 61070080, 61003184, 60832002; the National Grand Fundamental Research 973 Program of China under Grant No.2009CB320906; 863 Program of China under Grant No.2011AA01A110

## 5 References

1. S. Baker, T. Kanade, Limits on Super-Resolution and How to Break Them, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2000, pp. 372 - 379.
2. X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 35 (3) (2005)
3. H. Chang, D.-Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2004, pp. I-275 - I-282.
4. X. Ma, J. Zhang, and C. Qi, "Position-based face hallucination method," in Proc. ICME., pp.290 - 293, 2009.
5. X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," Pattern Recognition, vol.43, no.6, pp.3178 - 3194, 2010.
6. T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in Mathematical Models of Computer Vision, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York: Springer-Verlag, 2005.
7. Yang, J. C., J. Wright, et al. . "Image Super-Resolution Via Sparse Representation." IEEE Transactions on Image Processing 19(11): pp. 2861-2873.2010
8. M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing thenon-local-means to super-resolution reconstruction," IEEE Trans. ImageProcess., vol. 18, no. 1, pp. 36-51, 2009.
9. A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms with a new one", Multiscale Modeling and Simulation, Vol. 4, pp. 490-530, 2005.
10. Salmon, J. (2010). "On Two Parameters for Denoising With Non-Local Means." Ieee Signal Processing Letters 17(3): 269-272.
11. W. Gao, B. Cao, S Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. TSMC (Part A), 38 (2008), 149-161