IEICE TRANS. INF. & SYST., VOL.E95–D, NO.3 MARCH 2012

LETTER

Super-Resolution for Facial Images Based on Local Similarity Preserving

Jin-Ping HE[†a)], ***Member***, **Guang-Da SU**[†], ***and*** **Jian-Sheng CHEN**[†], ***Nonmembers***

SUMMARY To reconstruct low-resolution facial photographs which are in focus and without motion blur, a novel algorithm based on local similarity preserving is proposed. It is based on the theories of local manifold learning. The innovations of the new method include mixing point-based entropy and Euclidian distance to search for the nearest points, adding point-to-patch degradation model to restrict the linear weights and compensating the fusing patch to keep energy coherence. The compensation reduces the algorithm dependence on training sets and keeps the luminance of reconstruction constant. Experiments show that our method can effectively reconstruct 16×12 images with the magnification of 8×8 and the 32×24 facial photographs in focus and without motion blur.

key words: *super-resolution for facial images, manifold learning, local similarity preserving, point-based entropy, energy-coherent compensation*

1. Introduction

Super-resolution (SR) is a technology of reconstructing high-resolution (HR) images from low-resolution (LR) images according to certain prior knowledge. The traditional technique of SR has two categories: interpolation and reconstruction with sequences. Most of the restrictions in interpolating algorithms lead to the loss of high frequency (HF) information. On the other hand, the method of sequences-based reconstruction can not satisfy the requirement of large-scale arrangement in reality, because this method has high demands on collection equipment [1]. In recent years, a novel learning-based SR algorithm was proposed. It can not only overcome the limit of magnification and multi-frame of sequences-based reconstruction, but also nonlinearly add HF through learning manner. It has become an active research direction of SR in international area.

The method of learning-based SR is firstly proposed by Freeman et al. in 1999 [2]. They adopt Markov network to describe the relationship between LR patches and HR patches. Baker et al. in Carnegie-Mellon laboratory [3] proposed hallucinating faces in the same year. The 16×12 faces were well reconstructed with magnification of 8×8. Furthermore, this technology was applied on two facial photographs in focus and without motion blur [4], and one with good luminance was well reconstructed.

According to the scope of processing images, learning-based SR can be categorized into three classes: integrating global and local algorithm, global algorithm [5] and local algorithm [2]–[4]. In recent years, the new proposed approaches of learning-based SR [6]–[8] for facial images mainly tend to integrating global and local algorithm or local algorithm. The basic reason lies in the fact that the global algorithms just find the global linear structure of high dimension (HD) database. However, the local algorithm can obtain the nonlinear geometry structure of database. From the experimental results of existing articles, most testing images were acquired through down-sampling, and the best result was reconstructing 16×12 LR faces with the magnification of 8×8. There are just few actual results for face photographs. Besides Baker's [4], Zhuang et al. in Zhejiang University [9] and Liu et al. [10] show some reconstructed face photographs of their own.

In a word, firstly, integrating global and local algorithm or local algorithm has become the main development direction because of the nonlinear advantage; Secondly, the actual applications of face SR are very little, and there are just few results for face photographs in focus and without motion blur. Because the assumption of Locally Linear Embedding (LLE) is not always correct, a novel local SR algorithm based on manifold learning is proposed with directly restricting HR weights. Experimental results verify the effectiveness of our method.

2. Manifold Assumption and Advantages in SR

Manifold learning in general contains linear projection technologies, such as Principal Component Analysis (PCA), Multidimensional Scaling (MDS), and new nonlinear methods, such as ISOmetrical MAPing (ISOMAP), LLE, Laplacian Eigenmap, Local Tangent Space Alignment (LTSA) and so on. Nevertheless, linear methods of manifold learning just find the global linear structure of HD database, and they are incapable of reconstructing nonlinear geometry structure. Chang [11] in 2004 applied LLE to SR reconstruction. The algorithm implemented nonlinear estimation of the global image through linearly embedding local similar manifolds in HD space. Its basic assumption was that if a linear combination were expected to be constructed between each data point and its neighbors in HD space with a weights vector, each weight in the vector would be retained in low dimension space for each HD counterpart. Unfortunately, this is not always the case. As shown in Fig. 1 (a) it is a 3×3 patch being selected from one target LR face (size of 64×48 pixels). Its five neighbors are searched from one hundred sam-

Manuscript received May 6, 2011.
Manuscript revised November 2, 2011.

[†]The authors are with the Department of Electronic Engineering, Tsinghua University, 100084 China.
a) E-mail: hjping@mail.tsinghua.edu.cn

DOI: 10.1587/transinf.E95.D.892

ples (seen Fig. 1 (b)). The patch in Fig. 1 (a) can be represented by linearly combining the five neighbors in Fig. 1 (b) according to the method in [11]. The weights vector is $[0.82436\ 1.1322\ -0.59864\ -0.13612\ -0.19324]^{\mathrm{T}}$. On the other hand, the HR combining weights of 12×12 counterparts (seen Fig. 1 (c) and (d)) can also be calculated. The HR vector is $[0.35149\ 0.69683\ -0.31911\ 0.21261\ 0.096682]^{\mathrm{T}}$. From the above experimental results, each corresponding weights of the two vectors are completely different. The assumption of LLE which is keeping neighbors' zooming invariable is not always correct in SR applications.

However, in Fig. 1 it is worthwhile to note that after the target patch and its neighbors of LR are mapped to HR space, their counterparts of HR have the extreme similarity with each other. This is the applied foundation of local manifold learning in SR. It is identical with the assumption of Laplacian Eigenmap of nonlinear manifold learning. We name it "Local Similarity Preserving (LSP)" for short. In brief, the points of being close to each other in LR space were mapped into HR space, and the counterparts of HR will be also near by.

Furthermore, the local manifold learning methods can reconstruct the locally nonlinear geometry structure (seen Fig. 2). The center points of LR patches in Fig. 2 (a) and (c) are corresponding to the black rectangles of HR patches in Fig. 2 (b) and (d). The target HR patch of black rectangle contains some nonlinear HF. We can utilize the HF of black rectangles in Fig. 2 (d) to reconstruct the patch of Fig. 2 (b) through local learning. By contrast, the traditional interpolation impossibly restores the HF of black rectangle in Fig. 2 (b).

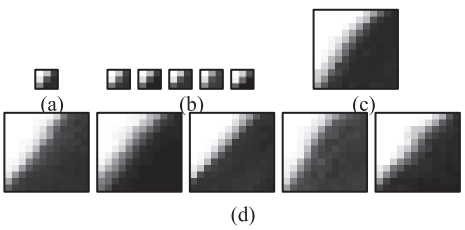

Fig. 1 An illustration of local similarity preserving with 4×4 magnification: (a) one input LR patch; (b) five nearest LR patches selecting from 100 training images; (c) the ground true HR patch corresponding to (a); (d) the HR patches corresponding to (b).

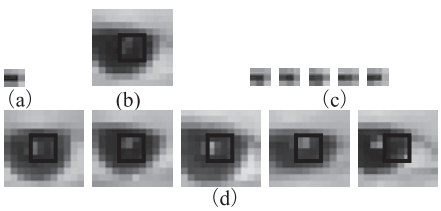

Fig. 2 The LSP advantage for reconstructing the locally nonlinear geometry structure: (a) one input LR patch; (b) the ground true HR patch corresponding to (a); (c) five nearest LR patches selecting from 10 training samples; (d) five HR patches corresponding to (c).

3. The SR Reconstruction Based on LSP

For solving the problem that the LLE assumption is not always correct, a new method which is directly restricting the HR weights is proposed. The inputs of the algorithm are the target LR image $\tilde{y}$ and HR training set $\{x_k\}_{k=1}^{T}$. Firstly faces $\{x_k\}_{k=1}^{T}$ were aligned according to the positions of two eyes and chin. We summarize our algorithm as follows:

Step 1: Acquire the LR training set according to degradation model;

Chang [11] utilized LLE to reconstruct the LR images without adding any priors, and the assumption excessively interfered with reconstructing effect. In this letter we adopt the degradation model [12] shown in (1) as prior.

$$\tilde{y}(i, j) = D(P_{ij}^{R}(h * \tilde{x})) + \eta \tag{1}$$

Here, h is the blurring function. $*$ represents convolution operation. $\tilde{x}$ is the HR image. η represents random noise. R is the magnification. $P_{ij}^{R}(\bullet)$ represents the image patch which row number is from $R(i-1)+1$ to Ri and column number is from $R(j-1)+1$ to Rj. $D(\bullet)$ is the sampling function. $\tilde{y}(i, j)$ is the gray intensity of point (i, j) for $\tilde{y}$. Our aim is reconstructing LR facial photographs in focus and without motion blur, so we set h to identity matrix and η to 0. $D(\bullet)$ is set to the mathematical function of image patch for the expected value.

The main task of first step is obtaining the magnification R according to the size of $\tilde{y}$ and $\{x_k\}_{k=1}^{T}$, and calculating LR training set $\{y_k\}_{k=1}^{T}$ according to (1).

Step 2: For one LR point $y(i, j)$, search its K nearest neighbors in the set $\{y_k(i, j)\}_{k=1}^{T}$ by mixing point-based entropy (PE) and Euclidian distance;

Baker [3] used the parent structure vector (PSV) as learning basis. The PSV consists of Laplacian pyramid, the horizontal and vertical derivative pyramid, and the second derivative pyramids. As shown in Fig. 3, the edge image obtained from the average of PSV does not contain enough HF in comparison with our proposed PE. The local features which PSV can provide are considerably limited. It can not embody all local details. Furthermore, the edge image after PE transformation has the abundant local textures (seen Fig. 3 (b)). So we can find the best matching points for $y(i, j)$ by the comprehensive parameter of mixing PE and Euclidian distance.

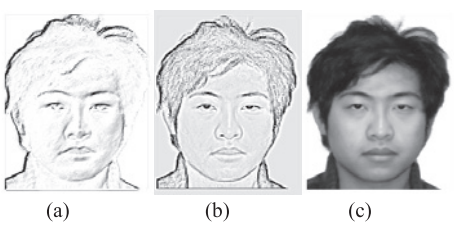

Fig. 3 Comparison of edge images using different methods: (a) the average edge image of Baker's PSV [3]; (b) edge images after PE transformation; (c) the original image.

According to the definition and characters of local entropy [13], we construct PE transformation. LR $\tilde{y}$ and $\{y_k\}_{k=1}^{T}$ are converted to PE feature space according to (2) and (3). Then we can obtain the corresponding entropy image $\tilde{H}$ and $\{H_k\}_{k=1}^{T}$.

$$\tilde{H}(i,j) = -C_{ij}\log_2(C_{ij}) \tag{2}$$

$$C_{ij} = \tilde{y}(i,j) \Big/ \sum_{m=-Z1}^{Z1}\sum_{n=-Z1}^{Z1} \tilde{y}(i+m, j+n) \tag{3}$$

Due to the point-to-patch degradation model, in LR space we adopt point values instead of patches. However, the point-to-point distance in LR space is terribly incorrect for measuring the similarity. So we select their neighbor windows as local geometry features for matching. The neighbor windows center at (i, j). The LR samples $\{\hat{y}_m\}_{m=1}^{K}$ corresponding to the K nearest points can be calculated by (4).

$$\begin{aligned}\{\hat{y}_m\}_{m=1}^{K} = \arg\min_{y_k} & \\ \times\Bigg(\alpha \sum_{p=-Z2}^{Z2}\sum_{q=-Z2}^{Z2} &\|\tilde{y}(i+p, j+q) - y_k(i+p, j+q)\|^2 \\ +\beta \sum_{p=-Z2}^{Z2}\sum_{q=-Z2}^{Z2} &\|\tilde{H}(i+p, j+q) - H_k(i+p, j+q)\|^2\Bigg)\end{aligned} \tag{4}$$

Record the HR $\{\hat{x}_m\}_{m=1}^{K}$ which are corresponding to $\{\hat{y}_m\}_{m=1}^{K}$, and the patches $P_{ij}^{R}(\{\hat{x}_m\}_{m=1}^{K})$.

Step 3: Linearly combine the HR patches $P_{ij}^{R}(\{\hat{x}_m\}_{m=1}^{K})$;

Suppose that $\{w_m\}_{m=1}^{K}$ represents the combination weights of the corresponding $P_{ij}^{R}(\{\hat{x}_m\}_{m=1}^{K})$ and $\tilde{x}'$ is the preliminary estimation of $\tilde{y}$. We can acquire the HR preliminary estimation of point $\tilde{y}(i, j)$ according to (5). The $\{w_m\}_{m=1}^{K}$ can be calculated through minimizing the sampling error $\|D(P_{ij}^{R}(\tilde{x}')) - D(P_{ij}^{R}(\tilde{x}))\|^2$, and it can be expressed by (6) referring (1).

$$P_{ij}^{R}(\tilde{x}') = \sum_{m=1}^{K} w_m P_{pq}^{R}(\hat{x}_m) \tag{5}$$

$$\{w_m\}_{m=1}^{K} = \arg\min \|D(P_{ij}^{R}(\tilde{x}')) - \tilde{y}(i,j)\|^2 \tag{6}$$

It is a constrained least squares problem for the unknown $\{w_m\}_{m=1}^{K}$. We can solve it through a linear equation system which is constructed by letting the first partial derivative of each w_m equal zero in (6).

Step 4: Implement energy-coherent compensation to linear combination patches;

When retrieving the training set $\{y_k(i, j)\}_{k=1}^{T}$ and the found patches are not very close to $y(i, j)$, the algorithm will produce the error fusing patch. For example, when the training set $\{y_k\}_{k=1}^{T}$ are all with the white background and the target $y(i, j)$ is at the gray background, it will be difficult to find the closest gray intensity at point (i, j) in $\{y_k(i, j)\}_{k=1}^{T}$. Apparently, the sampling point value represents a statistical energy of the corresponding HR patch. In order to preserve the energy coherence before and after reconstruction, and decrease the dependence of training set, we propose to compensate the sampling error instead of the LLE constraint $\sum_{m=1}^{K} w_m = 1$. The sampling error is calculated by (7) referring to (1).

$$e = D(P_{ij}^{R}(\tilde{x})) - D(P_{ij}^{R}(\tilde{x}')) = \tilde{y}(i,j) - D(P_{ij}^{R}(\tilde{x}')) \tag{7}$$

The final estimation of HR patch is obtained by (8).

$$P_{ij}^{R}(\tilde{x}) = P_{ij}^{R}(\tilde{x}') + e \tag{8}$$

This step makes each intensity of the preliminary patch increase e. Consequently, the expected value of the final patch also increases e. The found patches at the white background are converted to gray background through this self-adapted compensation. And this operation enhances the applicability of different databases.

Step 5: Smooth the blocking effect between adjacent HR patches;

The whole reconstructing face is obtained by mosaic of final HR patches. So there will be abundant of artificial blocking effects between adjacent patches. In order to reduce the artificial traces, our method is designed to reserve an overlap of one pixel between adjacent HR patches. We adopt a simple method to smooth blocking effect by averaging the gray intensities in overlapped regions between adjacent HR patches.

Step 6: Repeat Step2 to5 until all points of $\tilde{y}$ are calculated;

Step 7: Output the final HR estimation $\tilde{x}$.

4. Experiments

In order to demonstrate our method's efficacy and its applicability for different face database, we respectively perform our SR method on FERET database and Tsinghua University students' face database. Our method has five parameters to determine. The optimal parameters are all acquired through experiments. We set $Z1 = 2$, $Z2 = 1$, $\alpha = 0.9975$, $\beta = 0.0025$, $K = 5$ in the following results.

Firstly, FERET is considered as public database. The standard database contains more than 1000 images. Most of faces in these images are around 128×96 pixels or larger. The LR test faces are acquired through the degradation model shown in (1). We utilize a subset (549 images) of FERET for evaluation because more samples need a lot of time for registration. From Fig. 4 we can see that the results of our method are with fewer artifacts and more credible HF. The method in [5] just can reconstruct the 32×24 pixels while the SR result of LR 16×12 is seriously distorted. So in the sequent display for LR 16×12 the results in [5] are not listed.

The faces in FERET database are with different races, expressions and illumination conditions. To reduce the effect of the skin color, expression and illumination, we adopt the Tsinghua University students' face database as the training set for photographs in focus and without motion blur.

Suppose that the size of LR image $\tilde{y}$ is $M \times N$. The

Fig. 4 SR results of LR 32×24, 16×12 form top to bottom respectively with 4×4 and 8×8 manification. (a) LR images; (b) Bi-linear interpolation; (c) method in [5]; (d) method in [11]; (e) our results; (f) HR images.

Table 1 Comparison of RMS using different methods.

	Bi-linear	LLE in [11]	Our method
mean RMS of 100 test set	19.0368	18.0209	12.6066
RMS of Fig.5 female	47.2151	27.8600	17.6646
RMS of Fig.5 male	44.4184	18.5422	15.5530

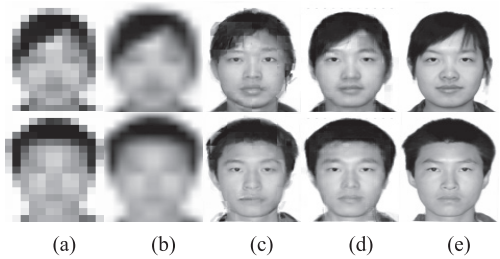

Fig. 5 SR results of LR 16×12 for 8×8 manification. (a) LR images; (b) Bi-linear interpolation; (c) method in [11]; (d) our results; (e) HR images.

objective assessment for reconstructed HR image $\tilde{x}$ is evaluated by Root Mean Square (RMS) which can be calculated by (9). We select the first 100 images in Tsinghua database as training set and select the following 100 images as test set. As shown in Table 1, the fist line show the mean RMS of 100test set after reconstructing LR 32×24 faces with $R = 4$. We can see that our method's RMS is obviously lower than others.

$$E = \sqrt{\sum_{i=1}^{RM} \sum_{j=1}^{RN} (\tilde{x}(i,j) - x(i,j))^2 / (R^2 MN)} \tag{9}$$

Figure 5 shows the reconstructed results of LR 16×12 using Tsinghua database. The other parameters are respectively $R = 8$ and $T = 500$. And our method outperforms other methods in global contour and details.

Furthermore, in Fig. 6 we show the photographs' SR results of LR 32×24 to display the reconstruction ability of our method. We set $T = 500$ and utilize Tsinghua database. From all the experiment results, we can see our method is effective and stable on different LR image sizes and different face database.

In this paper, a new method based on manifold learning is proposed. The important steps which make our method superior to the one in [11] are directly restricting the HR weights and further compensating the error instead of the conventional ways which make HR weights equal to LR weights and make the sum of HR weights equal to 1. The actual photographs can be successfully reconstructed. If we want to utilize this technology to reconstruct the actual suspects' faces, there are many researches to do yet, such as automatic LR face detection, face alignment, blur identification and training samples' selection.

Fig. 6 Photographs' SR results of LR 32×24: (a) LR images; (b) Bi-linear interpolation; (c) method in [11]; (d) our results.

Acknowledgments

The work is supported by National 973 plans projects of China (2007CB310600) and the National Science Foundation for Post-doctoral Scientists of China (20100470295)..

References

[1] J. Pu, J. Zhang, and H. Huang, "A survey of super-resolution algorithms," J. Shandong University, vol.39, no.1, pp.27–32, 2009.
[2] W.T. Freeman and E.C. Pasztor, "Learning low-level vision," ICCV, pp.1182–1189, Corfu. Greece, 1999.
[3] S. Baker and T. Kanade, "Hallucinating faces," Pennsylvania, Pittsburg: The Robotics Institute, pp.1–52, Carnegie Mellon University Press, 1999.
[4] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.9, pp.1167–1183, 2002.
[5] X. Wang and X, Tang, "Hallucinating face by eigentransformation," IEEE Trans. Syst. Man Cybern. C, Appl. Rev., vol.35, pp.425–434, 2005.
[6] B. Li, H. Chang, S. Shan, et al., "Aligning coupled manifolds for face hallucination," IEEE SPL, vol.16, no.11, pp.957–960, 2009.
[7] H. Huang, H. He, X. Fan, et al., "Super-resolution of human face image using canonical correlation analysis," Pattern Recognit., vol.43, no.7, pp.2532–2543, 2010.
[8] Y. Hu, K.M. Lam, G. Qiu, et al., "From local pixel structure to global image super-resolution: A new face hallucination framework," IEEE Trans. Image Process., vol.20, no.2, pp.433–445, 2010.
[9] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation," Pattern Recognit., vol.40, pp.3178–3194, 2007.
[10] C. Liu, H.Y. Shum, and W.T. Freeman, "Face hallucination: Theory and practice," IJCV, vol.75, no.1, pp.115–134, 2007.
[11] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," Proc. IEEE Conf. on CVPR, pp.275–282, 2004.
[12] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,"

IEEE Trans. Image Process., vol.6, no.12, pp.1646–1658, 1997.

[13] G. Wang, J. Tian, and J. Liu, "Infrared small objects detection based on entropy method," Infrared and Laser Engineering, vol.29, no.4, pp.26–29, 2000.