

Research on Face Restoring in Surveillance Videos

He Peng and Dong Wen

*College of Telecommunication and Electronic Engineering, Qiqihar University,
Qiqihar 161006, China
paper503@163.com, Dongwen917@163.com*

Abstract

Since cases being analyzed and detected by the police using surveillance videos, insufficient resolution and degradation always lead to the suspect's face being unrecognized. For such cases, from the perspective of learning-based face super-resolution method, this paper proposed a novel face restored solution based on face super-resolution and image mosaic. Firstly, by applying global similarity selection over the entire database, the scope of the used samples can be narrowed and boost the speed. Then the local representation between neighbors can make the method more robust against noise under severe conditions. Finally the reconstructed face could be inserted into original background by the improved Poisson Editing algorithm. This scheme achieve the effect of face restoring on a single frame in surveillance videos. Experiments under common database and real surveillance video scenarios demonstrate our method has better performance, especially for noise input images.

Keywords: *face super-resolution, global similarity selection, local similarity representation, position-patch, surveillance videos, Poisson Editing*

1. Introduction

In recent years, with the rise of high-resolution monitor system and intelligent analysis applications, public security departments using surveillance videos to detect, analysis and investigate cases has gradually become a trend. But when severe weather appears or the interested subject is too far from the camera, the most concerned area such as faces or license plates are still hard to identify. Therefore, an algorithm which could repair or rebuild an unrecognized target in surveillance videos will be of a great significance on the current society. This paper will research this technology and manage to carry out an implementation from the perspective of image super-resolution and mosaic.

Previous face restoring methods may use the face feature dictionaries to replace the blurred or incomplete face according to similarity [1-2]. But these methods are all depending on accurate locating and detecting of face landmarks, which is a difficult task under surveillance videos. Since then the research on face hallucination become a hotspot, which provide us a different way to rethink this problem. Freeman *et al.* [3] firstly used a patch based Markov network to represent the local geometry relationship between low-resolution (LR) and high-resolution (HR) patches, which costs lots of computation. Later, inspired by local linear embedding theory, Chang *et al.* [4] proposed Neighbor Embedding method to predict HR patch, but the unchanged number of neighbor patches will always lead to unstable reconstruction results. Ma *et al.* [5] took the fixed structure of human face as a prior and proposed a patch-position based face hallucination method, but if the number of training faces is larger than the LR patch's dimension, unstable results will appear due to the least square estimation. Then Yang *et al.* [4] introduced sparse representation to solved this problem, but the SR process is easily affected by noise and unknown degradations.

In surveillance videos, the original background often contains a lot of information. Therefore, the hallucinated HR face could be inserted into its background to provide a more complete information display result. And Poisson Editing [12] will be a useful method to solve this problem.

In this paper, by proposing a new face hallucination algorithm and improving the existing Poisson Editing method, a complete face restoring solution for surveillance video is formed.

2. Face Restoring Algorithm

A brief flow chart of the face restoring algorithm proposed in this paper is showed in Figure 1, which contains these main parts: Face Correcting and Extraction, Face Hallucination and Face Image Mosaic. This method is summarized as the following steps:

Step 1: The initial input image is a monitor frame including the target face. Eyes position is needed to be demarcated manually by the user, then the correction and extraction process will rotate the frame properly and extract the target face, and also obtain the original background.

Step 2: The extracted blurred face is then compared in the training set to select a small amount of samples, according to global selection algorithm based on 2D-PCA. These selected sample pairs will be used for reconstruction in the next step.

Step 3: Divide the target face and the selected sample pairs into patches, then the local representation algorithm is applied to reconstruct each patch according to its neighbor patches in the samples, and obtain the hallucinated HR face.

Step 4: Compute residual face between the hallucinated HR face and the blurred LR face. Take the residual face as a mask to guide the Poisson Editing process in the next step.

Step 5: Under the guidance of residual mask, apply Poisson Editing to insert the hallucinated HR face into the original background seamlessly.

Due to the face hallucination method proposed in this paper is a patch-position based method, therefore all the training faces and the input face are needed to be demarcated and calibrated by uniform standards.

For training samples, we use CAS-PEAL-R1 database which contains 1040 front expressionless faces. These training samples will be calibrated by the standard showed in Figure 2(a), eyes are firstly located manually and then the face correction process will extract the face according to the proportion between parameters Left, Right, Distance, Top and Bottom. For faces in surveillance videos, since this method is currently only for frontal face, and taking into account the possible angle offset, the calibration process is showed in Figure 2(b). The image will be firstly rotated until eyes are in a horizontal position, and then the extraction is the same procedure as above.

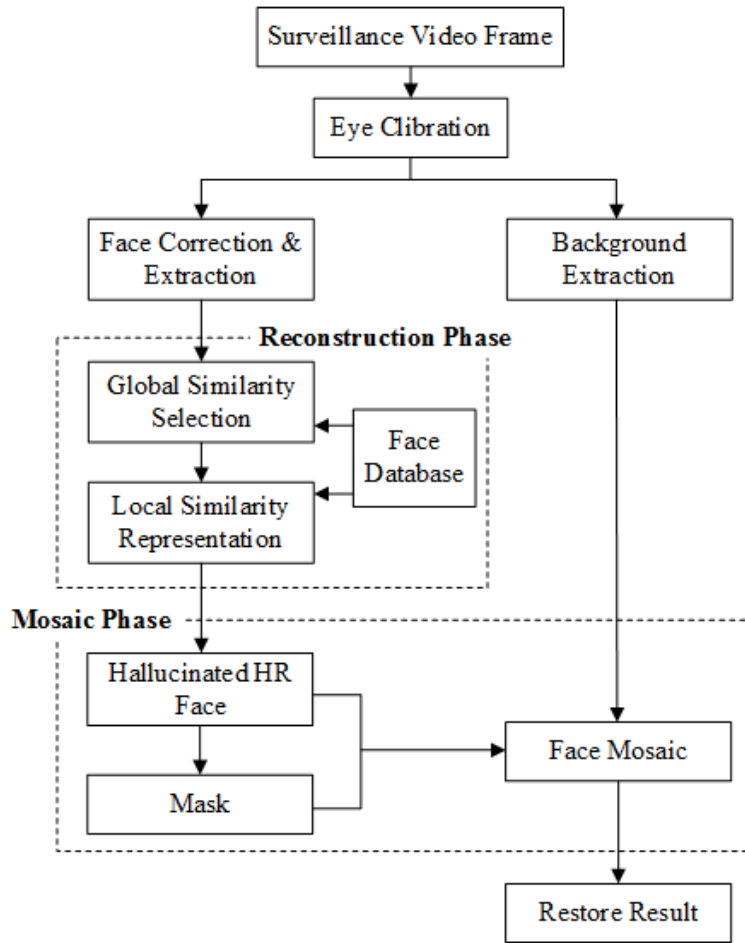


Figure 1. Flow Chart of Face Restoring Algorithm

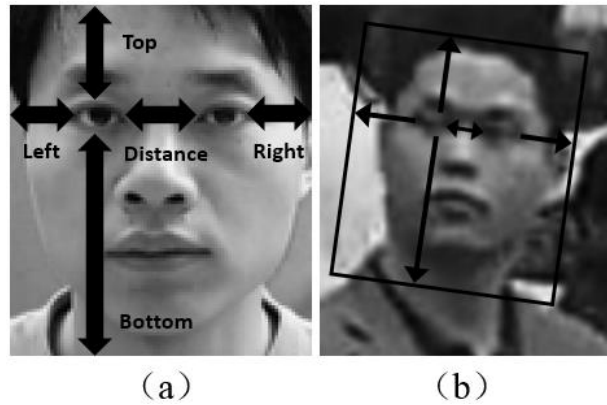


Figure 2. Face Correction and Extraction

3. Face Hallucination via Similar Selection and Representation

3.1. Formulation

Let N be the dimension of a LR patch and M be the number of samples in training set. Given the input LR patch $l \in R^{N \times 1}$, all the LR patches in training set which are at the same position as l can be represented as $L \in R^{N \times M}$, where the m th column $L_m(m=1, \dots, M)$

being the LR patch in sample m . Then the target LR patch can be represented as:

$$l = Lw + e \quad (1)$$

Where $w \in R^{M \times 1}$ is the weight vector and e is the reconstruction error vector.

Obviously the key issue of face hallucination is to solve the coefficients correctly. Constrained least square estimation could be used [3] as:

$$w^* = \arg \min_w \|l - Lw\|_2^2 \quad s.t. \sum_{m=1}^M w_m = 1 \quad (2)$$

But when the training set capacity M is much larger than the LR patch dimension, solution of the equation above becomes unstable. Then the sparse representation is introduced [6] into face hallucination and convert equation (2) into a standard SR problem:

$$\min_w \|w\|_0 \quad s.t. \|l - Lw\|_2^2 \leq \varepsilon \quad (3)$$

Where ℓ_0 norm represent the number of non-zero entries of w , ε is the error tolerance. And similar constrain based SR method [8] could be represented as:

$$w^* = \arg \min_w \{ \|l - Lw\|_2^2 + \lambda \|Dw\|_q \} \quad (4)$$

Where $D \in R^{M \times M}$ is the diagonal matrix that controls the constraint placed on the coefficients. After obtained the final reconstruction weights, the hallucinated HR patch could be computed as:

$$l_H = Hw \quad (5)$$

3.2. Global Similarity Selection

Traditional SR based face hallucination methods always use lots of training samples to represent the input data sparsely. Although the input data is polluted by noise and become not sparse at all, this process of representation will still costs heavy computation. According to the sparse theory, only a small portion of samples in the training set is useful to the reconstruction, so this procedure that spend a lot of time to compute the low value sample's coefficients is obviously unwise.

Therefore, for the images corrupted by noise or which under unknown degradation such as surveillance videos, we no longer seek for the coefficients that can represent data's sparsity but use similarity to describe the input data directly. In order to exclude the worthless samples in the perspective of similarity, we introduce the 2D Principle Component Analysis (2D-PCA) theory to achieve this process, as showed in Figure 3.

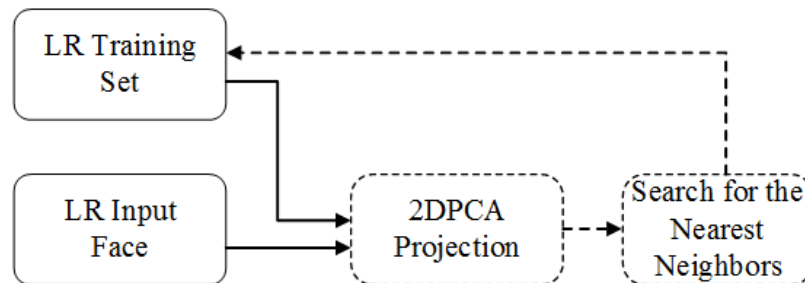


Figure 3. Global Similarity Selection

For the projection of LR training set, covariance matrix is firstly computed as:

$$G_t = \frac{1}{M} \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \quad (6)$$

Where \bar{A} is mean face and $A_j(j=1, \dots, M)$ is each training sample. Then the eigenvectors correspond to the largest d eigenvalues are extracted and form the projection matrix $X=\{X_1, \dots, X_d\}$. And project each sample in the training set into the projection space as:

$$Y_m = L_m X, m=1, 2, \dots, M \quad (7)$$

Where L_m is the sample m . After completing the projection of training set, projection of input face L_{input} is the same as above:

$$Y_{input} = L_{input} X \quad (8)$$

Finally, by computing the Euclidean distance $\|Y_m - Y_{input}\|_2^2$ between L_{input} and Y_m , we can select the most similar samples compared to the input face from the entire training set.

But how many similar samples should be picked out is obviously a crucial problem. Figure 4 show how the number of similar samples affect the quality of reconstruction. We randomly choose 800 faces from the database for training and the other 240 for testing, the average PSNR and SSIM values of reconstruction results are displayed in Figure 4. The horizontal axis represents how many samples are cut out, for example 5% means there are 5% of samples in the database are excluded. We can see the reconstruction quality is basically unchanged until abscissa is larger than 60%, which means cutting off some samples properly will not affect the reconstruction quality, but save lots of time. Therefore, in the following tests we set the number of samples needed to be selected to 40% of the database.

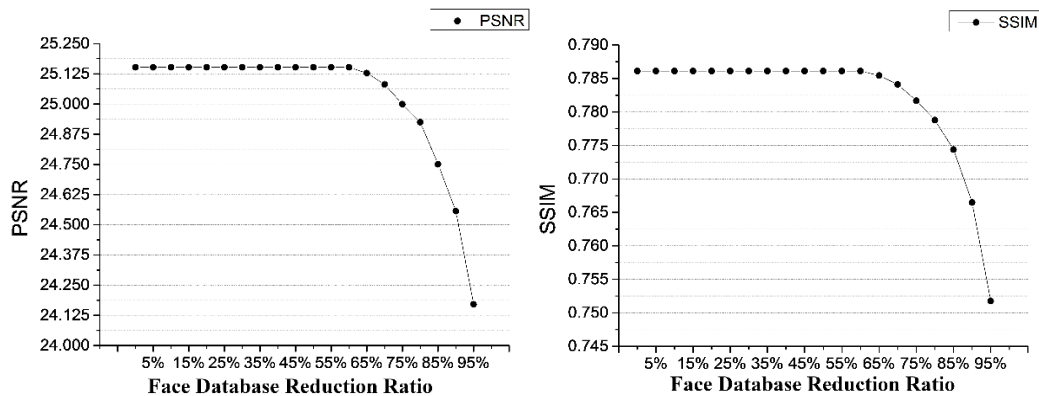


Figure 4. Average PSNR and SSIM with the Number of the Similar Faces Changes

3.3. Local Similarity Representation

After picking out the similar faces from the entire set. To make full use of information in the neighbor patches, we propose a parallel two layer structure to represent patch, which is showed in Figure 5. The patches marked by dark solid lines in each sample mean that they are the neighbor patch whose coefficient should be calculated during the representation. Because every sample have a same procedure to represent patch, so this method has a natural parallel structure which could be implemented paralleled easily.

Let C be the number of selected similar faces in 3.2 and S be the number of neighbor patches considered in representation ($S=9$ in Figure 5). Given an input LR patch $l \in R^{N \times 1}$, and supposed that all the neighbor patches in m th sample are represented as $L_m^s \in R^{N \times S}$, which the s th column is L_m^s ($s=1, \dots, S$). Then the reconstruction for LR patch l is as following steps:

Step 1: For each sample from the 1st to the M th, let coefficients for patch 1 be $w_m \in R^{M \times 1}$, which represents the similarity between its neighbors and itself. And w_m is computed as :

$$w_m^s = \exp\left(-\|l - L_m^s\|_2\right), 1 \leq s \leq S, 1 \leq m \leq M \quad (9)$$

Step 2: After obtained w_m , compute the middle-layer HR patch for each sample:

$$h_{H1m} = \frac{1}{G_m} H_m^S w_m, G_m = \sum_{s=1}^S w_m^s, 1 \leq m \leq M \quad (10)$$

Where H_m^S represents the HR neighbor patches set corresponding to the m th sample, and G_m is the summation of entries in w_m .

Step 3: Since all M middle-layer HR patch h_{H1m} ($m=1, 2 \dots M$) have been computed, the final HR patch h_{H2} can be synthesized as:

$$h_{H2} = \frac{1}{\sum_{m=1}^M G_m} \sum_{m=1}^M G_m h_{H1m} \quad (11)$$

Finally, by assembling all hallucinated patches according to their position and averaging the overlapped pixels, the estimated HR face is obtained.

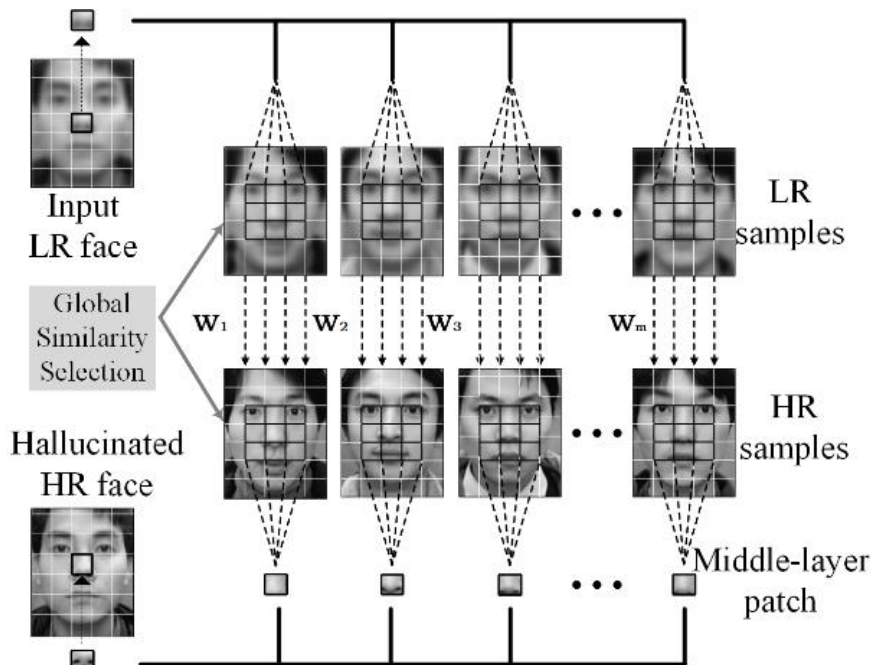


Figure 5. Local Similarity Representation

4. Face Mosaic Based on Poisson Equation

Poisson edit [12] can provide seamlessly mosaic or edit on specific image area. It's main idea is solving the Poisson differential equation with Dirichlet boundary conditions to get the pixel values in the mosaic area. Let g be source image, f^* be destination image, and the result be f . The mosaic area is Ω and its boundary is $\partial\Omega$. Poisson equation can be written as:

$$\begin{cases} \Delta f|_{\Omega} = \Delta g|_{\Omega} \\ f|_{\partial\Omega} = g|_{\partial\Omega} \end{cases} \quad (12)$$

Where Δ is Laplacian operation. This equation guarantee that the pixels are the same inside and outside the boundary, while the result have the same texture information as the source image.

In practical applications, a mask is used to instruct the mosaic area Ω . But if we set the mask's shape the same as the HR face's, once big reconstruction errors such as dark or light spots occurred on the boundary, the Dirichlet boundary condition will wrong which lead to the wrong solution of Poisson equation. To overcome this problem, we extend the boundary of mosaic area spread all over the face. This will greatly reduce the possibility of errors on the boundary leading to the wrong equation, thus all the details will be saved in mosaic process. We need to get the residual face firstly: $diff = I_H - I_{input}$, and the mask is the binarization of it:

$$\begin{cases} Mask(diff \geq thresh \times mean(diff)) = 1 \\ Mask(diff < thresh \times mean(diff)) = 0 \end{cases} \quad (13)$$

Where $thresh$ is an empirical parameter. By taking account the mask, source image and background image into the Poisson equation, the solution could be solved and get the result. An instance of the mosaic process described above is showed in Figure 6.



(a) Input Face (b) HR Face (c) Mask

Figure 6. An Example of Face Mosaic

5. Experiment and Results

We conduct experiments on the proposed face hallucination algorithm to testify its performance under different level of noises and real scenarios. We also test the performance of the entire face restoring method using some randomly selected face in a surveillance video. NE [4], LcR [7] methods are also tested for better comparison.

The database we applied for training is the CAS-PEAL-R1 Chinese face database, which contains 1040 frontal faces. We randomly select 800 faces for training and the other 240 faces for testing. All the sample faces are cropped to the same size 100×84 . The smooth kernel is fixed by a 8×8 Gaussian low-pass filter with a deviation of 4. All the sample faces are being degraded by smooth kernel and then down-sampled with factor

4. So the size of LR sample is 25×21 . The patch size is fixed by 3×3 with overlap of 1 pixel.

We add a zero-mean Gaussian noise with different noise levels ($\sigma=0,2,4,6,8$) on the input face. And the hallucination results when the noise level is 4 is showed in Figure 7. We can see that our proposed method which is based on similarity selection and representation is able to resist noise and synthesize a more smooth and detailed face. The corresponding average PSNR and SSIM values are listed in Table 1 and Table 2. Relative to the second best method LcR, the PSNR gains of our proposed method is about 3dB, and the SSIM gains is nearly 0.1.

For the tests under real scenarios, we randomly cropped 5 faces in a street surveillance video, and the reconstruction results are showed in Figure 8. We can see that on the basis of good integrity and similarity, the proposed face hallucination method recovered more detailed information than NE and LcR method. Finally, the entire face restoring scheme is tested and the 4 pairs of face mosaic results are showed in Figure 9. The original partial frame of the video is on the left side and the right side is the corresponding restoring result. According to the restoring results, the improved Poisson editing basically could insert the HR face seamlessly into background, and the extra information recovered by our proposed face hallucination method is well integrated with the original information. So this scheme could be a convenient information retrieval and display method for professional use.



(a) Bicubic (b) NE (c) LcR (d) Proposed (e) Original

Figure 7. Hallucination Results of Different Methods ($\sigma=4$)

Table 1. PSNR Comparison of Different Methods Under Gaussian Noise (Db)

Noise level	NE	LcR	Proposed
$\sigma=0$	27.21	27.68	30.74
$\sigma=2$	26.74	25.93	29.51
$\sigma=4$	25.11	25.38	28.25
$\sigma=6$	24.63	24.27	26.92
$\sigma=8$	23.24	23.10	25.87

Table 2. SSIM Comparison of Different Methods Under Gaussian Noise

Noise level	NE	LcR	Proposed
$\sigma=0$	0.7324	0.7690	0.8671
$\sigma=2$	0.7271	0.7574	0.8513
$\sigma=4$	0.7117	0.7092	0.8112
$\sigma=6$	0.6899	0.6501	0.7862
$\sigma=8$	0.6597	0.6711	0.7591



(a) Bicubic (b) NE (c) LcE (d) Proposed

Figure 8. Hallucination Results of Different Methods Under Real Scenarios



Figure 9. Surveillance Face Restoring Examples

6. Conclusion

In this paper, aiming at the problem that faces in low resolution videos are always unrecognizable, we proposed a face restoring method which consists of hallucination phase and mosaic phase. By applying global similarity selection and local similarity representation, our reconstruction algorithm has a lower time complexity and is able to resist noise under real scenarios. By computing residual face as mosaic mask, we improved Poisson editing so that more smooth mosaic results are obtained. Experiments under face database and real scenarios also demonstrate the effectiveness and feasibility of this method.

Acknowledgements

This work was partly supported by the research of science and technology education department of HeiLongjiang Province with grand NO. 2013GS230301.

References

- [1] T. Bourlai, A. Ross and A. K. Jain, "Restoring Degraded Face Images: A Case Study in Matching Faxed, Printed, and Scanned Photos", *Journal IEEE Trans. Information Forensics and Security*, vol. 6, no. 2, (2011), pp. 371-384.
- [2] T. Bourlai, A. D. Clark and L. S. B. Rowden, "Methodological insights on restoring face photos of multinational passports. Technologies for Homeland Security (HST)", *IEEE International Conference, Waltham, MA*, (2013).
- [3] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision", *International Journal Computer. Vis.*, vol. 40, no. 1, (2000), pp. 25-47.
- [4] H. Chang, D. Y. Yeung and Y. Xiong, "Super-resolution through neighbor embedding. *Computer Vision and Pattern Recognition*", *Proceedings of the 2004 IEEE Computer Society Conference, Washington, DC, USA*, (2004).
- [5] X. Ma, J. Zhang and C. Qi, "Position-based face hallucination method", *Multimedia and Expo. IEEE International Conference, New York, USA*, (2009).
- [6] J. Yang, H. Tang, Y. Ma, and T. Huang, "Image super-resolution via sparse representation", *J. IEEE Trans. Image Process.*, vol. 19, no. 11, (2010), pp. 2861-2873.
- [7] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation", *J. IEEE Trans. Multimedia*, vol. 16, (2014), pp. 1268-1281.
- [8] Z. Wang, J. Jiang, Z. Xiong, R. Hu, and Z. Shao, "Face hallucination via weighted sparse representation", *J. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, (2013), pp. 2198-2201.

- [9] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face hallucination via weighted adaptive sparse regularization", J. IEEE Trans. Circuits and Systems for Video Technology, vol. 24, (2013), pp. 802-813.
- [10] J. Zhang, C. Zhao and R. Xiong, "Image super-resolution via dual-dictionary learning and sparse representation", J. In: Proc. IEEE Int. Symp. Circuits Syst., (2012), pp. 1688-1691.
- [11] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces", Computer Vision and Pattern Recognition, IEEE Computer Society Conference, Maui, HI, (1991).
- [12] H. Wu and D. Xu, "Improved Poisson image editing and its implementation on GPU", Computer-Aided Industrial Design & Conceptual Design, IEEE 10th International Conference, Wenzhou, China, (2009).
- [13] H. M. M. Naleer, Y. Lu and H. M. M. Naleer, "A new two-step face hallucination through block of coefficients", Computer Science and Automation Engineering, IEEE International Conference, Zhangjiajie, China, (2012).
- [14] C. Liu, H. Y. Shum and C. S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model", Computer Vision and Pattern Recognition. Proceedings of the 2001 IEEE Computer Society Conference, (2001).

Authors



He Peng, (1970), Ph.D. Professor and Master Tutor of Qiqihar University. His current research interests including image processing, intelligent control and information technology.



Dong Wen, (1990), Received B.S. degree in Nanjing University of Posts & Telecommunications, Nanjing, China, in 2013. Currently working toward M.S. degree at Qiqihar University. Research interests include image processing and super-resolution.

