International Journal of Multimedia and Ubiquitous Engineering
Vol.8, No.6 (2013), pp.47-58
http://dx.doi.org/10.14257/ijmue.2013.8.6.05

Super-Resolution for Surveillance Facial Images via Shape Prior and Residue Compensation

Tao Lu[1,2], Ruimin Hu[2], Junjun Jiang[2],Yanduo Zhang[1] and Wei He[3]

[1]Hubei Province Key Laboratory of Intelligent Robot, College of Computer Science and Engineering Wuhan Institute of Technology, Wuhan, 430073, China
[2]National Engineering Research Center on Multimedia Software, Wuhan University, Wuhan 430072, China
[3]Transportation Engineering Institute of Minjiang University, Fuzhou 350001, China

lut@whu.edu.cn, {hrm1964, junjun0595}@163.com, zhangyanduo@hotmail.com

Abstract

Super-resolution (SR) methods are widely used to enhance the resolution of input images such as low-resolution (LR) remote or surveillance images. The key problem of SR algorithm is to introduce a proper prior for high-resolution (HR) image reconstruction. Usually pixel feature based similarity measure is used as prior. While in real surveillance applications, images are often disturbed by noise so that it is not enough to only use the prior from pixel domain to achieve satisfied high-resolution. In this paper, a two-phase based face hallucination approach via shape prior (SP) is proposed. Firstly, the Active Appearance Model (AAM) is used as shape prior to reconstruct the global face. Then the high frequency information as detailed inartificial facial features of HR face image is compensated to the global face. Experiments show that the proposed algorithm improves the subjective and objective quality of the input LR facial images and outperform many states-of-the- art super-resolution methods.

Keywords: *Facial image, Super-resolution, Shape prior, Global face, two-step method*

1. Introduction

Super-resolution (SR) is an effective tool in the area of criminal investigation, surveillance, remote sensing [1]. Especially in a surveillance system, it is very difficult for identification and recognition due to the low-resolution (LR) human facial images. Therefore, in order to get detailed facial features for recognition, it is necessary to infer a potential high-resolution (HR) image from input LR ones, this technique is called face hallucination. In general, face hallucination techniques can be classified into two categories: reconstruction-based approaches and learning-based approaches. Reconstruction-based methods utilize input multi-LR-images via sub-pixel alignment to fuse one HR output image. These methods depend on the accuracy of alignment. While in surveillance situation, high-accuracy alignment of LR images is very hard as the low quality images and local object motion. So these methods only have limited performance in real surveillance scene. Recent years, motivated by machine learning theory, leaning based methods use sample data to infer the missing high-frequency detail information. As the useful information from the samples, learning based methods have satisfied subjective and objective image quality. Learning based methods become more and more popular and a hot research issue. In this paper we focus on the learning based hallucination methods in real surveillance scene.

ISSN: 1975-0080 IJMUE

In recent years, numerous learning-based face hallucination methods are proposed. These methods infer the HR face image from a training dataset which comprises a series of pairs of HR and LR face image samples. Baker *et al*., [2] firstly proposed a learning based approach method by facial pixel feature prior. Gaussian pyramid was used to model the relationship of HR and LR image. When inputting LR patch, the nearest neighbors in LR sample space would be found to match the inputs, then corresponding HR image or patch in Gaussian pyramid are tread as HR outputs. Multi-resolution pyramid model was useful to represent the relationship of HR and LR images, even this model was build up, it was not robust to noise or disturbance in inputs. This method treated the image as patches, so called as Local face super-resolution algorithm. Not as common image, facial images have fixed structure, such as eye, nose and mouth, so the facial structure information was used to optimize the output HR facial images. Wang and Tang [3] proposed a new global hallucination method using subspace learning. Principle Component Analysis (PCA) was used to fit the input face image as a linear combination of the LR images in the training set. And then the same weights were used to synthesis the inferred HR images. This methods treat one face image as a whole to infer HR image, so this method also called as Global face super-resolution algorithm. The facial images were decomposed into different components by PCA algorithm. Then high and low frequency information were separated. So this method was robust to noise, but there are ghost effects at the edges of face due to the over-fit of samples. To overcome this problem, Liu *et al*., [4] proposed a two-step method combining global and local face method to get a better output HR image. Global face method was used to reconstruct an initial result in the first step, and then local face method was utilized to compensate the missing high frequency parts. This two-step method takes advantage of both global and local face approach, comparing to either global or local methods, has better performance. Inspired by this method, two-step frameworks are popular used to achieve better results in many papers [4, 5]. Huang [5] proposed a two-step method approach based on Canonical Correlation Analysis (CCA), and CCA was used separately both for global face appearance and local details. Not like PCA algorithm, correlation of HR and LR image was represented by CCA which had been proved beneficial to SR reconstruction. In the same time, another local patch-based approach has been proposed. Chang *et al.*, [6] first introduced manifold learning theory to model the SR problem. Yang *et al.*, [7] proposed a sparse representation based local patch method. These patch-based methods have smooth output but not robust to noise. All these methods achieve good results in simulation with pixel feature prior, but in the case of real surveillance situation, the input images are always damaged by random noise and blur, the pixel feature are also damaged so it is hard to achieve satisfied results in noise or blur situation.

In order to resolve face SR problem in real surveillance situation, Gajjar *et al*., [8] proposed a learning based super-resolution method by inhomogeneous Gaussian Markov random field prior. The database acquired by real camera, which can reveal the real degradation process with real prior. This method utilize frequency domain to model the multi-resolution relationship. When the input image is suffered by noise, the frequency representation error may increase. So in noise situation, this method can't obtain satisfied results, and this SR method deals with common image not for facial images. For real surveillance situation, Lan *et al*., [9] proposed a semantic face shape constrains to global face reconstruction for real low-quality face images, the shape prior are used to represent the intrinsic feature of facial images. This method is robust to noise but the outputs still have ghost effect on edges. And the outputs facial image lacks of detail information. To overcome this problem, we proposed a novel face shape prior in SR as first step, and use the second step to compensate the residuals as high-frequency detailed information. Compared to [9], we proposed a two-step framework by simultaneously minimizing the HR image and shape parameters with alternating iteration and compensating the residue into the final results. As shown in Figure.1.

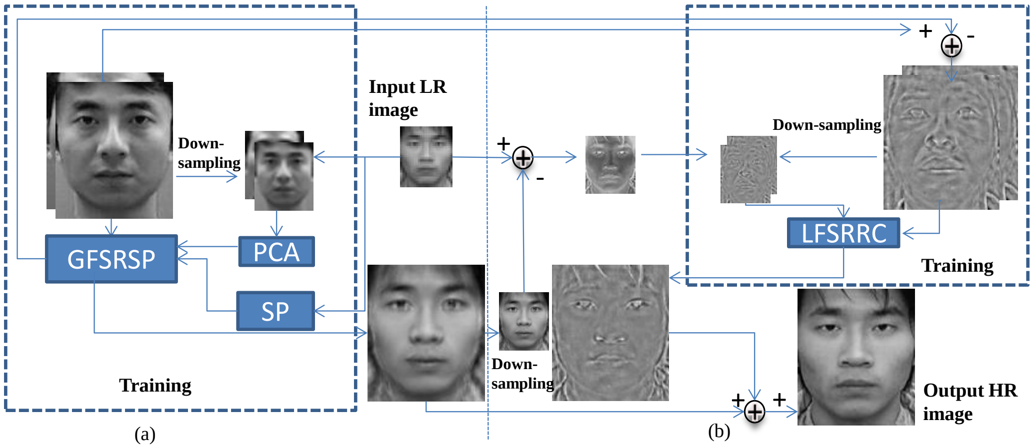

Figure 1. Architecture of the proposed two-step SR algorithm. (a) Global face reconstruction; (b) Residual compensation.

In this paper, we proposed a novel two-step SR framework as illustrated in Figure 1. In the first step, a novel global faces SR via shape prior (GFSRSP) is proposed as global face method. PCA models are built for LR and HR facial images as global face reconstruction, and shape prior (SP) regularization term makes the output face shape similar to the input shape. The detail of algorithm will be described in Section 3. In order to recover further high-frequency inartificial facial features, we proposed a local face SR by residue compensation (LFSRRC) using neighbor embedding to recover HR residual face in the second step. And then, the final HR facial image is obtained by adding the first HR global facial image and the second residual facial image. The paper is organized as following: in Section 2 point distribution model is used to describe the face shape. And in Section 3, face SR via shape prior and residual compensation is introduced. Section 4 is about experiment results of the proposed algorithm. And the last section is the conclusion.

2. Face Shape Prior by Point Distribution Model

Park *et al.*, [10] proposed an example-based face hallucination method by morphable face model. The face is separated into an extended shape and extended texture. Experiment results verify that the face shape and texture are important for HR reconstruction, and shape information is more robust to noise than texture information. So in this paper, we use point distribution model to describe the face shape.

2.1 Face shape model

Point distribution model is widely used to locate the shape of an object. Cootes *et al.*, [11] proposed active appearance model (AAM), which represents the location of shape point. For a calibrated image, the shape vector can be calibrated by each feature point coordinates, the coordinates of all feature points arranged to form shape vector x:

$$x = (lx_1, ly_1, lx_2, ly_2, \cdots, lx_m, ly_m)^T \tag{1}$$

Where the coordinates of a shape feature point is (lx, ly), m indicates the number of shape characteristic points. AAM can decouple the face shape and texture, if the shape points are calibrated, then each training image is warped to match those of the mean shape to obtain a shape-free patch, which is raster scanned into a texture vector g . We normalize it by a linear transformation. The correlations between shape and texture are learned to get a combined appearance model.

$$x = \bar{x} + Q_s b \quad \text{and} \quad g = \bar{g} + Q_g b . \tag{2}$$

Here, $\bar{x}$ is the mean face shape, $\bar{g}$ is the mean texture in a mean-shape patch, Q_s and Q_g are matrices describing the modes of variation derived from the training set. b is the appearance model parameters, which can control the face shape and texture.

In this paper, we define the face shape with 48 landmarks shown in Figure 2. These 48 landmarks represent specific points of faces. If inputting a facial image, the task is to locate the landmarks which contain the face shape information. The appearance parameter b is modeled by AAM and can locate the shape landmarks automatically.

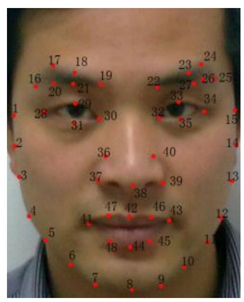

Figure 2. Face shape feature points

2.2 Shape parameters search algorithm

While the face shape model is built up by AAM. The most important step is to resolve the appearance parameters by iteration. Given an input face image g_{im} , after wrapping texture into mean shape and normalization by transformation T_u^{-1}, then the input image is represented as $g_s = T_u^{-1}(g_{im})$. The current model texture is $g_m = \bar{g} + Q_g b$. Then the difference between model and image is defined as:

$$f(b) = \arg\min \| g_s - g_m \|^2 = \arg\min_b \| T_u^{-1}(g_{im}) - \bar{g} - Q_g b \|^2 \tag{3}$$

Minimizing the equation (3), we can get the least Square solution. Here parameters b containing two parameters which represent face shape and texture, in this method we focus on the shape parameter. In this algorithm, the training and testing data are aligned and normalized to the mean face shape. Different from the purpose of AAM algorithm to locate the landmarks, the purpose of our method is to get the appearance parameter which can restrict the output image have similar feature as the input. Then the face shape points can be used as prior in the HR image reconstruction which will be discussed in the next section. As we know, the accuracy of shape parameter may affect the role of prior in HR reconstruction. Especially, for the LR facial image, it is hard to locate the shape points precisely, so it is necessary to revise the coordinates by hands. So the shape point's coordinates are got by semi-automated algorithm and hands.

3. Two-step Framework of Face Super-resolution by Shape Prior

Once, for an input LR image, we can get the face shape points as prior to reconstruct the output HR image. Especially in the global face reconstruction, the shape prior can be used to reconstruct a HR image which has similar shape with the input.

3.1 Global face SR by shape prior

Suppose the input low-resolution image is y, the observed high-resolution image is H, D is the down-sampling matrix, B is the blur matrix, n is the noise matrix, then we have:

$$y = DBH + n \tag{4}$$

SR is a typical inverse problem. Equation (4) is underdetermined, there are many solutions, and so different kinds of image prior are used to get a satisfied solution. In this paper, we add the face shape prior into the cost function to constraint the output image. And the cost function is:

$$\arg(H,b) = \arg\min_{H,b}(\| y - DBH \|^2 + \lambda_1 \| \Gamma(H) \|^2 + \lambda_2 \| R(b) - b_0 \|^2) \tag{5}$$

The first term is to constraint the output image to look like the observed low-resolution one, the second term is image smooth prior, and the last one is face shape prior to keep the shape consistent between inputs and outputs, λ_1 and λ_2 are balance parameters to balance the image shape, image smooth prior and image texture, function $R(b)$ is the above procedure to get the appearance parameter which is the inverse function of formula(2) , b_0 is the input face shape parameter.

To solve the energy-minimization in Equation (5), the alternating minimization Algorithm is used to get the high-resolution image and face shape parameters. Firstly suppose $b = b_0$, get the solution of H_n, then let $R = B^T D^T DB + \lambda_1 \Gamma\Gamma^T + \lambda_2 ff^T$ and $P = B^T D^T y$ with $H_{n+1} = H_n - \mu(RH_n - P)$. Secondly let $H = H_n$, get the solution of b_n, then let $r = Q_q^T T_u^{-1}(H) - Q_q^T \overline{g} - Q_q^T b_0$ and $p = Q_g^T Q_g$ with $b_{n+1} = b_n - \beta(rb_n - p)$,here μ and β are the steps of steepest descent method. H and b are both alternating minimized until convergence, algorithm details can be found in [12].

Followed by above optimization framework, we add the shape prior into the global face SR algorithm. PCA is used to represent the LR and HR image respectively. Suppose a set of face images represent by an $N \times M$ matrix, $[l_1, \ldots l_M]$, where l_i is the image vector, N is the number of image pixels, M is the number of samples. Removing the mean face m^t from each image, we have $L = [l_1 - m^t, \ldots l_M - m^t]$ as the matrix to represent the sample data. The eigen faces are computed from the eignen vectors of ensemble covariance matrix: LL^T, where T is transpose. Decomposing the matrix $L^T L$ by SVD, then the eigen faces will be as:

$$E = LV_t \Lambda_t^{-\frac{1}{2}} \tag{6}$$

Where V_t and Λ_t are respectively eigenvector matrix and eigenvalue matrix of $L^T L$. For input image y vector, a weight vector can be computed by projecting in onto the eigen face:

$$w^g = E^T(y - m^t) \tag{7}$$

As we know, in LR and HR subspace the weight vector projecting onto LR and HR eigen faces matrix respectively cannot satisfy the manifold consistency assumption. So the weight vectors need to transform to sample subspace which the LR and HR images which is considered have manifold consistency. We have a new weight vector:

$$c = V_t \Lambda_t^{-\frac{1}{2}} w^g \tag{8}$$

Then input image y can be expressed by $y = Lc + m^t$. In [], the weight vector c is supposed the same in LR and HR subspaces. So the potential HR image can be expressed by:

$$H = L_H c + m_H^t \tag{9}$$

Where L_H , c and m_H^t are respectively represent HR sample data removing mean face, LR subspace weight vector and HR mean face. Then formula (5) can be rewritten as:

$$\arg(c,b) = \arg\min_{c,b}(\| y - DB(L_H c + m_H^t)\|^2 + \lambda_1 \| \Gamma(L_H c + m_H^t)\|^2 + \lambda_2 \| R(b) - b_0 \|^2) \tag{10}$$

In order to solve this minimization problem, alternating minimization Algorithm is used to get the high-resolution image weight vector and face shape parameters simultaneously. The output HR images will be treated as the first step result by global face SR algorithm.

3.2 Local face SR by residue compensation

The global high-resolution face image got from above steps looks smooth but loses some detailed feature. To overcome this problem, we add the high-frequency image information often called "residue compensation" onto the global face result. In this paper, the face residue on a face is divided into patches, and then local neighbor patch based method is used to map the low residue patch to the high residue one. Suppose I^l , I^h are LR input and high-resolution output, I^o is the original high-resolution face. Then LR residue is $R^l = I^l - D(I^h)$, high-resolution residue face is $R^h = I^o - I^h$.

Let r^m represent the LR residue, m is the number of training face images, P is the number of patches. Each face residue is divided into patches as $\{r^{mP}(i,j)\}_{P=1}^N$, and their reconstruction weight matrix $w^l(i,j)$, which represents the contribution of each training image patch located at position (i,j) to the reconstruction of the input face images that have the same positions. Here N represents the number of overlapped patches in each face.

Let y^r indicate the input testing face residue. Using the training dataset patches to represent the input face images. The reconstruction error is:

$$\arg(w^l) = \arg\min_{w^l} \| y^P(i,j) - \sum_{m=1}^{M} w^l(i,j) r^{mP}(i,j) \|^2 \tag{11}$$

To minimize Eq (11), the least squares method can used to get the reconstruction weights, then keeping the weights to the high-resolution residue patches, the high-resolution residue r can be synthesized.

As shown in Figure 1, once the input LR image, we can get an initial HR global face image in the first step and a HR residue image in second step. So we add the residue image to the global face image as the latent HR output image. The algorithm will be summarized as follow:

Algorithm 1. SR for surveillance facial images via Shape Prior and Residue Compensation

First step:

Input: Training set L and L_H , LR input image y , input shape parameter b_0 .

Output: HR image H

Training Phase: Get LR eigenfaces matrix E , LR mean face m_L^t , HR mean face m_H^t , Q_s and Q_g ,mean shape $\bar{x}$,mean texture $\bar{g}$.

1. Compute initial weight vector of input image y by (6),(7),(8)
2. Use (10) to compute initial shape parameter b
3. Compute c by alternative minimization
4. Output HR image H by (9)

Second step:

Input: LR residue matrix r^m ,HR residue matrix r_H^m from the first step, input LR residue image y^r .

Output: HR residue image r .

Training Phase: LR residue matrix position patch matrix, HR residue matrix position patch matrix.

1. Compute each position input LR weight vector $w^l(i,j)$ by (11)
2. Keep the weight vector into HR residue matrix space get HR residue patch
3. Combine each patch to a HR residue image

Output: HR residue image r

Final output: $H+r$ as the output of this algorithm.

4. Experiment

4.1 Database and Parameters Setting

To verify the superiority of our method, experiments are conducted on Chinese face database CAS-PEAL-R1 [13], there are 1040 individuals with one front face image for each individual. We use 1000 images and calibrate the shape points by hand for training, the rest 40 images as test images. The high resolution image size is fixed at 96×112 pixels. All images are blurred and down-sampled to low resolution images of 24×28 pixels. For each input face images, we interpolate the input image size to be that of high-resolution ones and calibrate the input shape points as prior. In our method, the parameters are set by the empirical value, $\lambda_1=0.001$, $\lambda_2=0.0001$. In our framework, if λ_2 is zero, the method is the traditional face SR method, 98% principal components are preserved both in shape and texture representation. In the second step, patches' size is 20×20 pixels, and the overlap is 8 pixels. Both of HR image and shape parameters get stable solution after 20 iterations. All other face SR methods are set to the best performance in their paper.

4.2 Results

We compare our method with Lan's shape semantics method [9], Liu's two step method [4], and Huang's two step method [5] in the same database. 6 individuals are chosen for testing, the result are demonstrated in Figure 3. Different from Lan's algorithm, alternately iteration is used for more accurate solution in our algorithm, which has almost same subjective performance, such as (c) and (e). When we compensate residue to the global face, more face details are found in (f), so the residue face image which more high-frequency information is useful for face reconstruction. Although Huang's result looks like smoother and less edge aliasing in (d), the faces seems to similar with the mean face and lose detailed information. Different with Liu's method, our algorithm use position-based neighbor reconstruction method to compensate the residue instead of Markov random network, it is obvious to find that our algorithm achieves better subjective quality.

To quantify the deviations of the results from the ground truth data, we calculate the PSNR of each method, Gaussian noise is used to simulate the real noise with standard deviation of 0.001, and the results are listed in Figure 4. Compared to Lan's methods, our algorithm gains better performance because of error compensation. Compared to Huang's and Liu's methods, our method achieves better global face features by shape prior and iteration framework. The PSNR have higher value than all of other stat-of-art SR algorithms.

In order to test our method in real situation, we get a real world image from a surveillance camera, as shown in Figure 5. We get the surveillance image with red frame to locate the face by hands, and then get face shape feature points as prior; the size of LR input is 24×28 pixels. (d) is the result of cubic interpolation. (c) has apparent aliasing especially on the moth and eyes. (d) looks like the mean face. We observe that our methods have better subjective quality, in (f) edges of faces and detailed feature are clearer than others, especially the eyes have better quality, and this conclusion is consistent with simulation.

Figure 3. Results of different hallucination methods (a) The input low resolution image (b) Liu' global face method (c) Lan's method (d) Hunag's two step method (e) the first step of our algorithm (f) the proposed method (g) Original image

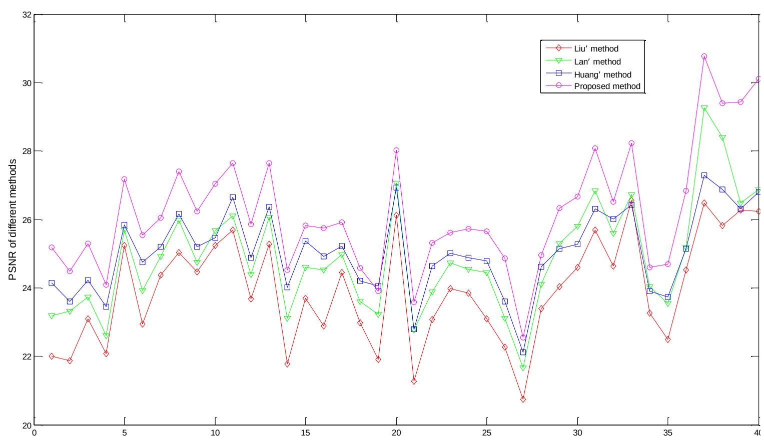

Figure 4. PSNR value of different super-resolution methods in noise condition, 40 images are used for test

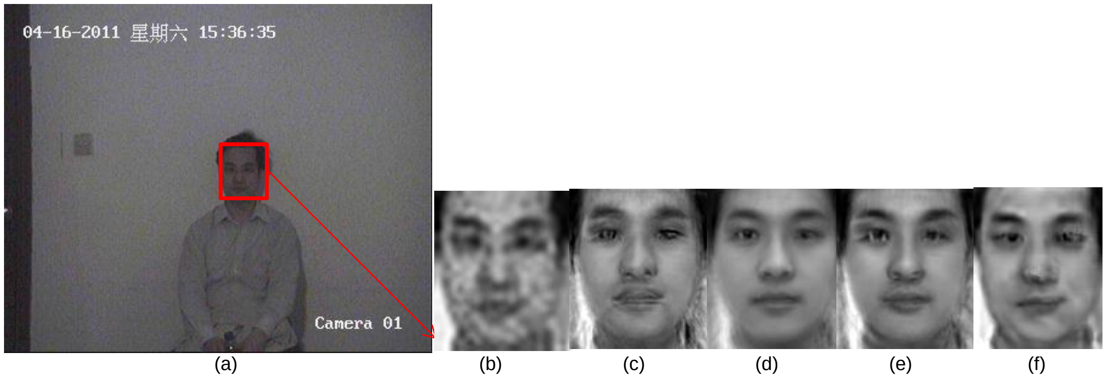

Figure 5. Real world image experiments. (a) real world input image (b)Cubic interpolation (c)Liu's method (d) Huang' method (e) Lan's method (f)the proposed method

5. Conclusion

In this paper, we proposed a novel two-step face SR method. In the first step, we use face shape as prior to constraint the global face SR, and in the second step we compensate the residue to the global face. Face shape feature points are used as reconstruction prior which makes the output images have similar shape with the inputs. At the same time, we use local patch-based method to compute the HR residue image and add it to the global face image. Compared to other state-of-art hallucination methods, our approach has two advantages: robustness to noise and more detailed face features. Subjective and objective results verify the effectiveness of this algorithm.

Acknowledgements

This research was supported by the National High-Tech Research and Development Program("863"Program,No:2013AA12A202),the National Natural Science Foundation of China (61172173,61070080,60970160,61003184), the Natural Science Foundation of Hubei Province of China (2012FFA099, 2012FFA134), the youths science foundation of Wuhan institute of technology, the excellent youth science and technology innovation team project of the Educational Department of Hubei Province of China (No. T201206, B2013219), Wuhan Planning Project of Science and Technology(2013010602010217), Project from Wuhan Information Industry bureau.

References

[1] J. Jiang, R. Hu, Z. Han and T. Lu, "Efficient single image super-resolution via graph-constrained least squares regression", Multimedia Tools and Applications, (**2013**), pp. 1-24.

[2] S. Baker and T. Kanade, "Hallucinating Faces", Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, (**2000**), pp. 83-88, Grenoble.

[3] X. Wang and X. Tang, "Hallucinating Face by Eigentransformation", IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications And Reviews, vol. 35, no. 3, (**2005**), pp. 425-434.

[4] L. Ce, Shum and H. Y. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (**2001**), Kauai, HI, USA.

[5] H. Huang, H. T. He, X. Fan and J. Zhang, "Super-resolution of human face image using canonical correlation analysis", Pattern Recognition, vol. 43, (**2010**), pp. 2532–2543.

[6] H. Chang, D. Yeung and Y. Xiong, "Super-resolution through Neighbor Embedding", Proceedings of IEEE Conference of Computer Vision and Pattern Recognition, (**2004**), pp. 275-282.

[7] J. Yang, J. Wright, T. Huang and Y. Ma, "Image superresolution via sparse representation", IEEE Transactions on Image Processing, vol. 19, no. 11, (**2010**), pp. 2861-2873.
[8] P. P. Gajjar and M. V. Joshi, "New Learning based super-resolution: Use of DWT and IGMRF prior", IEEE Trans. on Image Processing, vol. 19, no. 5, (**2010**), pp. 1201-1213.
[9] C. D. Lan and R. M. Hu, "A Face Super-Resolution Approach Using Shape Semantic Mode Regularization", Proceedings of the 17th IEEE International Conference on Image Processing, **(2010)**, pp. 2021-2024.
[10] J. -S. Park and S.-W. Lee, "An Example-Based Face Hallucination Method for Single-Frame, Low-Resolution Facial Images", IEEE Transactions on Image Processing, vol. 17, no. 10, (**2008**), pp. 1806-1816.
[11] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active Appearance Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, (**2001**), pp. 681-685.
[12] W. Gao, B. Cao and S. Shan, "The CASPEAL Large-scale Chinese face database and baseline evaluations", IEEE Trans on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 38, no. 1, (**2008**), pp. 149-161.
[13] D. Zachariah, M. Sundin, M. Jansson and S. Chatterjee, "Alternating Leat-Squares for Low-Rank Matrix Reconstruction", IEEE Signal Processing Letters, vol. 19, no. 4, (**2012**), pp. 231-234.

Authors

Tao Lu

He received the B.S and M.S degrees from Computer Science and Engineering Department, Wuhan institute of technology , Wuhan, China in 2003 and in 2009 respectively, and Ph.D. degree in Communication and Electronic System from Wuhan University in 2013. He is lecturer of Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of technology. His research interests include image/video processing, computer vision and artificial intelligence.

RuiMin Hu

He received the B.S and M.S degrees from Nanjing University of Posts and Telecommunications, Nanjing China, in 1984 and in 1990 respectively, and Ph.D. degree in Communication and Electronic System from Huazhong University of Science and Technology, in 1994. Now he is a professor and the director of National Engineering Research Center For Multimedia Software, Wuhan University and Key Laboratory of Multimedia Network Communication Engineering in Hubei province. He is Executive Chairman of the Audio Video coding Standard (AVS) workgroup of China in Audio Section. His research interests include audio/video coding and decoding, video surveillance and multimedia data processing.

