Aligning Coupled Manifolds for Face Hallucination

Bo Li, Hong Chang, *Member, IEEE*, Shiguang Shan, *Member, IEEE*, and Xilin Chen, *Member, IEEE*

***Abstract*—Many learning-based super-resolution methods are based on the manifold assumption, which claims that point-pairs from the low-resolution representation manifold (LRM) and the corresponding high-resolution representation manifold (HRM) possess similar local geometry. However, the manifold assumption does not hold well on the original coupled manifolds (i.e., LRM and HRM) due to the nonisometric one-to-multiple mappings from low-resolution (LR) image patches to high-resolution (HR) ones. To overcome this limitation, we propose a solution from the perspective of manifold alignment. In this context, we perform alignment by learning two explicit mappings which project the point-pairs from the original coupled manifolds into the embeddings of the common manifold (CM). For the task of SR reconstruction, we treat HRM as target manifold and employ the manifold regularization to guarantee that the local geometry of CM is more consistent with that of HRM than LRM is. After alignment, we carry out the SR reconstruction based on neighbor embedding between the new couple of the CM and the target HRM. Besides, we extend our method by aligning the multiple coupled subsets instead of the whole coupled manifolds to address the issue of the global nonlinearity. Experimental results on face image super-resolution verify the effectiveness of our method.**

***Index Terms*—Manifold alignment, super-resolution.**

I. Introduction

SUPER-RESOLUTION (SR) is to estimate a high-resolution (HR) image from a single or multiple given low-resolution (LR) counterpart(s). As a popular research topic in image processing and computer vision, various SR algorithms have been proposed during the past decades. The traditional interpolation-based technique is based on generic smoothness priors and usually results in blurring problem. Some edge preserving methods, such as [1], are proposed to address this problem. Another kind of commonly used methods is the reconstruction-based approach which is based on the principle that the down-sampled HR estimation should be close to the input LR image. In [7], Back-Projection (BP) algorithm is proposed to minimize the reconstruction error in an iterative manner. To avoid the artifacts in the reconstructed images, some works propose to use edge-directed priors (e.g., [13]).

Recently, learning-based SR methods have attracted more attention. Methods of this class can be dated back to the well-known early work proposed in [4] and [2]. Freeman *et al.* [4] use a patch-wise Markov network to model the relationship between the LR images and the HR counterparts. Baker and Kanade [2] develop a learning-based method named "face hallucination" for face image SR. Besides, Liu *et al.* [8] propose a two-step modeling approach that integrates global and local models. In [11], the authors introduce primal sketch priors (i.e., edges and corners) into the learning-based framework to further improve the performance of SR.

Inspired by a promising manifold learning method, locally linear embedding (LLE) [10], a series of learning-based SR approaches with a manifold assumption have been presented. This assumption states that the points on high-resolution representation manifold (HRM) share the same local geometry with their counterparts on low-resolution representation manifold (LRM). Here, the local geometry is characterized by the reconstruction weights of the points in their neighborhoods. Chang *et al.* [3] first use neighbor embedding (NE) method to solve SR problem and achieve fairly good performance. Then, another method in [5] is proposed to combine the advantages of NE with image primitive priors. More recently, authors of [14] introduce sparse representation into NE by adding a L_1-norm term into the reconstruction objective function.

Although NE and its following-up methods have achieved fairly good performance in SR, there are few discussions on whether the coupled manifolds, LRM and the corresponding HRM, really satisfy the manifold assumption. Authors of [5] only validate that the assumption is more consistent with primitive patches than general ones. In this letter, we revisit the manifold assumption and find that it does not hold well. The basic reason lies in the fact that the one-to-multiple mapping from HR to LR is nonisometric. Therefore, previous methods based on such assumption usually result in blurring and artifacts.

To avoid the limitations of previous works, we propose a new single image hallucination method based on manifold alignment. Different from the previous manifold alignment method, such as [6], we propose to learn two explicit mappings to align LRM and HRM into one common manifold (CM). Meanwhile, to achieve the goal of SR estimation, some HRM regularizations are employed to guarantee the local geometry of CM is more consistent with HRM than LRM is. The alignment task is finally formulated as an optimization problem which has an analytical solution. However, real-world LRM and HRM are usually too complex to be aligned well by only one couple of linear mappings. We thus extend our method to multiple subsets alignments to address this problem. We conduct experiments on publicly available face databases to verify the efficacy of the proposed algorithm and corroborate the above claims.

Manuscript received March 26, 2009; revised June 17, 2009. First published July 14, 2009; current version published August 26, 2009. This work was performed at Institute of Computing Technology (ICT) and supported by the NSFC under Contracts 60803084 and U0835005 and by the National Basic Research Program of China (973 Program) under Contract 2009CB320902. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhou Wang.

B. Li is with the School of Computer Science and Technology, Harbin Institute of Technology, China (e-mail: bli@jdl.ac.cn).

H. Chang, S. Shan, and X. Chen are with Key Laboratory of Intelligent Information Processing, ICT, Chinese Academy of Sciences (e-mail: hchang@jdl.ac.cn; sgshan@jdl.ac.cn; xlchen@jdl.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2009.2027657

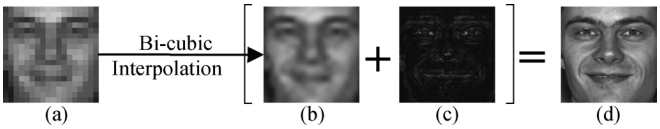

Fig. 1. Relationship among (a) LR image; (b) LF part; (c) missing HF part; and (d) HR image.

II. Manifold Assumption

A. Preliminaries

As shown in Fig. 1, we first perform an interpolation on an input LR image to predict the low-frequency (LF) part of HR target. Then we estimate the missing high-frequency (HF) part of the HR image and add it to the LF part to output the final HR estimation. In this context, our method only focuses on the prediction from LF to HF. From now on, we use the LF part as the representation of a LR image and the HF part as that of a HR image, which is similar with the representations in [4].

As previous algorithms, we need construct a patch-wise training set. At first, the LR images are generated from the training HR images by blurring and subsampling. Then the LF part is interpolated from each LR image. Immediately, the HF part is got by subtracting the LF part from the corresponding HR image. Finally, the LF parts and the HF parts are separated into two patch-wise LR and HR representations: $\{\mathbf{l}_i\}_{i=1}^{N} \subset \Re^m$ and $\{\mathbf{h}_i\}_{i=1}^{N} \subset \Re^M$. Here, i is the patch index and N is the number of patches. In general, these two patch sets are considered to be sampled from the LRM and HRM respectively. Furthermore, we use two matrices $\mathbf{L} = [\mathbf{l}_1\ \mathbf{l}_2 \cdots \mathbf{l}_N]$ of size $m \times N$ and $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_1 \cdots \mathbf{h}_N]$ of size $M \times N$ to represent the training set. For a test LR image, we compute its LR representation (i.e., LF part) and then split it into patches denoted as $\{\mathbf{l}_x\}_{x=1}^{K}$. x is the patch site index in the image plane and K is the number of patches in the test image. The target of our algorithm is to predict the corresponding patch-wise HF part, $\{\mathbf{h}_x\}_{x=1}^{K}$. We normalize the patches using the method in [4]. In next subsection, we will revisit the manifold assumption.

B. Revisiting Manifold Assumption

Generally, the neighbor embedding algorithm based on the manifold assumption could be illustrated in Fig. 2(a) and (b), which includes the following steps: 1) for each query LR patch representation $\mathbf{l}_x$, find its k-nearest neighbors on LRM; 2) calculate the weights $\mathbf{w}_x^L$ for reconstructing $\mathbf{l}_x$ by its neighbors $\mathbf{l}_j$; 3) transfer the neighbors and weights to the corresponding HR patch representations on HRM, and estimate the target HR counterpart $\mathbf{h}_x$ by a linear combination. The last two steps can be formulated using the following mathematical equations:

$$\arg\min_{\mathbf{w}_x^L} \|\mathbf{l}_x - \sum_{j \in N_L(x)} \mathbf{w}_{xj}^L \mathbf{l}_j\|^2 \ \text{s.t.} \ \sum_{j \in N_L(x)} \mathbf{w}_{xj}^L = 1 \tag{1}$$

and

$$\hat{\mathbf{h}}_x = \sum_{j \in N_L(x)} \mathbf{w}_{xj}^L \mathbf{h}_j. \tag{2}$$

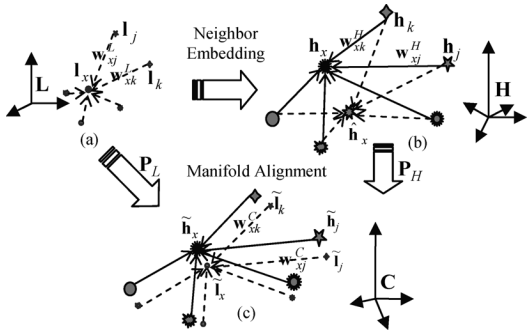

Fig. 2. Neighbor embedding based on manifold assumption and manifold alignment. $\tilde{\mathbf{x}}$ is the projection of $\mathbf{x}$ on CM.

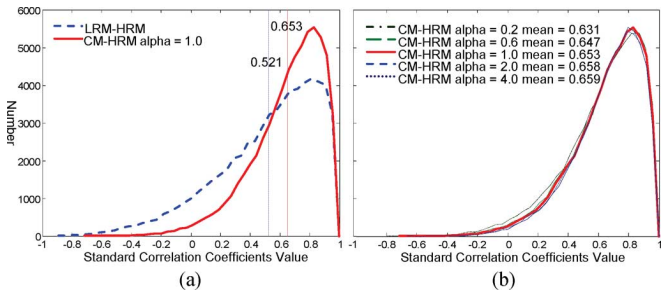

Fig. 3. Distributions of standard correlation coefficients between reconstruction weights on the couple manifolds of LRM-HRM, and CM-HRM with different α, respectively.

Here, $N_L(x)$ consists of the neighbor indexes of patch $\mathbf{l}_x$.

According to the manifold assumption, the reconstruction weights of one LR patch representation on LRM should be extremely similar with the weights of reconstructing the HR counterpart with the corresponding neighbors on HRM. Unfortunately, this is not always the case. In Fig. 3(a), we draw the distribution of the standard correlation coefficients between the reconstruction weights at the corresponding positions (i.e., $\mathbf{w}_x^L$ and $\mathbf{w}_x^H$) on LRM and HRM. We can see that the mean of the values is only 0.521. After aligning the couple of LRM and HRM to CM, as the case shown in Fig. 2(c), this distribution between $\mathbf{w}_x^C$ and $\mathbf{w}_x^H$ at each site x is also calculated on CM and HRM and centers at 0.653. It means that the couple of CM and HRM are more consistent with manifold assumption, and the alignment processing may improve the performance of the NE-based SR method.

III. Hallucination Based on Manifold Alignment

A. Formulation and Solution

In this letter, we propose a method to align LRM and HRM into CM by two linear projections. According to the SR target, the alignment should achieve two objectives. On one hand, the embeddings of each patch pair (i.e., one LR patch representation and its HR counterpart) should be as close as possible on CM, which is called as *correspondence preserving objective*. On the other hand, the local geometry (e.g., reconstruction weights) of the HRM should be preserved on the CM, which is called as *manifold regularization*. According to these principles, our minimization function of coupled manifold alignment method could be expressed as follows:

$$J(\mathbf{P}_L,\mathbf{P}_H)=\sum_{i=1}^{N}||{\mathbf{P}_L}^T\mathbf{l}_i-{\mathbf{P}_H}^T\mathbf{h}_i||^2 +\alpha\sum_{i=1}^{N}||{\mathbf{P}_H}^T\mathbf{h}_i-\sum_{j\in N_L(i)}\mathbf{w}_{ij}^H{\mathbf{P}_H}^T\mathbf{h}_j||^2. \quad (3)$$

Here, $\mathbf{P}_L$ represents the projection matrix from the points on LRM to the embeddings on CM. $\mathbf{P}_H$ represents the projection matrix from the points on HRM to the embeddings on CM. α is a harmonious parameter used to balance these two terms. In the objective function (3), the former term is the correspondence preserving term. The latter term is the manifold regularization term which is inspired by LLE [10]. Weight $\mathbf{w}_{ij}^H$ is calculated by reconstructing $\mathbf{h}_i$ using its neighbors $\{\mathbf{h}_j\}$.

For this optimization problem, we propose an analytical solution as follows. Firstly, let $\mathbf{P}=\begin{bmatrix}\mathbf{P}_L\\ \mathbf{P}_H\end{bmatrix}$, $\mathbf{Z}=\begin{bmatrix}\mathbf{L} & \\ & \mathbf{H}\end{bmatrix}$ and $\mathbf{A}=\begin{bmatrix}\mathbf{I} & -\mathbf{I}\\ -\mathbf{I} & \mathbf{I}+\alpha\mathbf{M}\end{bmatrix}$. Here $\mathbf{I}$ represents the identity matrix and

$$\mathbf{M}=(\mathbf{I}-\mathbf{W})^T(\mathbf{I}-\mathbf{W}),\mathbf{W}_{ij}=\begin{cases}\mathbf{w}_{ij}^H, & j\in N_L(i)\\ 0, & \text{otherwise.}\end{cases}$$

Then we transfer the objective function (3) into a new concise form as

$$J(\mathbf{P}_L,\mathbf{P}_H)=\mathrm{Tr}\left(\mathbf{P}^T\mathbf{Z}\mathbf{A}\mathbf{Z}^T\mathbf{P}\right) \quad (4)$$

where $\mathrm{Tr}(.)$ is a trace operator. To achieve scaling and translation invariance, we introduce the constraints and solve the optimization problem such as

$$\min_{\mathbf{P}_L,\mathbf{P}_H} J(\mathbf{P}_L,\mathbf{P}_H)\ \text{s.t.}\ \mathbf{P}^T\mathbf{Z}\mathbf{Z}^T\mathbf{P}=\mathbf{I}\ \text{and}\ \mathbf{P}^T\mathbf{Z}\mathbf{e}=0 \quad (5)$$

where $\mathbf{e}=[11\ldots1]^T$ is the vector of ones with $2N$ entries.

Furthermore, let $\mathbf{E}=\mathbf{Z}\mathbf{A}\mathbf{Z}^T$ and $\mathbf{F}=\mathbf{Z}\mathbf{Z}^T$, the solution to the optimization problem with respect to $\mathbf{P}$ could be given by the 2nd to $(d+1)$-st smallest generalized eigenvectors of $\mathbf{E}\mathbf{p}=\lambda\mathbf{F}\mathbf{p}$. After matrix $\mathbf{P}$ of size $(m+M)\times d$ is calculated, the coupled mapping matrices $\mathbf{P}_L$ of size $m\times d$ and $\mathbf{P}_H$ of size $M\times d$ could be constructed.

It is worth noting that $\mathbf{F}=\mathbf{Z}\mathbf{Z}^T$ usually is noninvertible. To avoid the singular case, we propose the regularization as $\mathbf{F}=\mathbf{Z}\mathbf{Z}^T+\tau\mathbf{I}$, where τ is set to a small positive value (e.g., $\tau=10^{-6}$).

B. Proposed Hallucination Algorithm

However, the real-world problem of face hallucination is usually too complex to align the coupled manifolds by only one pair of projection matrices. To address this problem, we carry out the multi-alignment on the subsets of the coupled manifolds, respectively. However, how to divide the manifolds into coupled subsets is also a problem. Fortunately, all face images have similar structures. The patches at the same site are highly related once we align the face images according to the positions of two eyes. In the light of this property, we divide the training sets into some coupled subsets according to the site x, which are represented by the coupled matrices $\mathbf{L}^x=[\mathbf{l}_1^x\,\mathbf{l}_2^x\cdots\mathbf{l}_{N_x}^x]$ and $\mathbf{H}^x=[\mathbf{h}_1^x\,\mathbf{h}_2^x\cdots\mathbf{h}_{N_x}^x]$, N_x denotes the number of patches in the subsets at site x. When the test image contains K patches, there are K coupled subsets in the training set. In the final scheme, the coupled mapping matrices $\mathbf{P}_L^x$ and $\mathbf{P}_H^x$ are learned at each site x. After aligning the manifolds, we perform the NE-based SR reconstruction between the CM and HRM. We summarize our algorithm in Table I.

TABLE I
NEIGHBOR EMBEDDING BASED ON MANIFOLD ANIGNMENT

Input: Training set $\{\mathbf{L}^x\}_{x=1}^K$ and $\{\mathbf{H}^x\}_{x=1}^K$, LR-patch representations $\{\mathbf{l}_x\}_{x=1}^K$.

Output: HR-patch representations $\{\mathbf{h}_x\}_{x=1}^K$.

Offline phase:

Learn the coupled projection matrices on the training subsets at each site x and obtain $\{\mathbf{P}_L^x\}_{x=1}^K$ and $\{\mathbf{P}_H^x\}_{x=1}^K$.

Online phase:

For each LR query representation $\mathbf{l}_x$,

1. Find its k nearest neighbors $\mathbf{l}_j^x$ in the subset $\mathbf{L}^x$ by Euclidean distance;
2. Project $\mathbf{l}_x$ and its neighbors $\{\mathbf{l}_j^x\}$ onto CM by $\mathbf{P}_L^x$ as follow:
$$\widetilde{\mathbf{l}}_x={\mathbf{P}_L^x}^T\mathbf{l}_x\ \text{and}\ \widetilde{\mathbf{l}}_j^x={\mathbf{P}_L^x}^T\mathbf{l}_j^x;$$
3. Calculate reconstruction weights by minimizing
$$J(\mathbf{w}_x^C)=\|\widetilde{\mathbf{l}}_x-\textstyle\sum_j\mathbf{w}_{xj}^C\widetilde{\mathbf{l}}_j^x\|^2+\beta\sum_y\|\widehat{\mathbf{h}}_y-\widehat{\mathbf{h}}_x\|^2\ \text{s.t.}\ \sum_j\mathbf{w}_{xj}^C=1. \quad (6)$$
Here, y is a x's adjacent site; $\widehat{\mathbf{h}}_y$ and $\widehat{\mathbf{h}}_x$ denote the overlap of adjacent patches $\mathbf{h}_y$ and $\mathbf{h}_x$, respectively where $\mathbf{h}_x=\sum_j\mathbf{w}_{xj}^C\mathbf{h}_j^x$;
4. Reconstruct HR patch representation by $\mathbf{h}_x=\sum_j\mathbf{w}_{xj}^C\mathbf{h}_j^x$;

Repeat Step 1 to 4 until all HR representations are estimated.

Finally, we mosaic all the estimations $\{\mathbf{h}_x\}_{x=1}^K$ to get the HF part and add it to the LF part to output the final HR image.

In Table 1, Step 3, we introduce a term $\sum_y||\widehat{\mathbf{h}}_y-\widehat{\mathbf{h}}_x||^2$ to smooth the adjacent patches in HR image plane. β is a balance parameter. Different from previous algorithms, in Steps 2 and 3 we calculate the reconstruction weights on the CM rather than the original LRM. In next section, the experimental results will verify that our algorithm is fairly effective.

IV. EXPERIMENTS

We perform experiments of the single face image SR on FERET [9] face Database to demonstrate the efficacy of the proposed method. All 72×72 HR facial images used in experiments are aligned with the positions of two eyes. The HR image is smoothed and subsampled to LR sizes of 24×24, 18×18, and 12×12 pixels, respectively. The standard training set contains 1002 frontal face images. We evaluate the method on a subset (773 images) of the *fafb* probe sets (1195 images), which is selected to avoid including any images from the subjects of the training set. We use 12×12 patch with overlap size of 4 pixels between adjacent patches in HF and LF image plane, and set $\beta=0.2$. The setting of the number of nearest neighbors is an open problem in LLE. In our experiment, we

Fig. 4. Face SR results of LR 12×12, 18×18, 24×24 are shown from top to bottom. (a) HR images; (b) results of Bi-cubic interpolation; (c) results of method in [4]; (d) results of method in [3] ; (e) results of method in [14]; (f) our results; (g) LR images.

Fig. 5. Face SR results of LR 18×18. (a) HR images; (b) results of bi-cubic interpolation; (c) results of method in [4]; (d) results of method in [3] ; (e) results of method in [14]; (f) our results; (g) LR images.

set $k = 5$, which is consistent with the setting of [3]. From Fig. 3(b) we can see that the distribution and the mean of the standard correlation coefficients are both stable for variant parameter α in (3). So, we use the trade-off setting of $\alpha = 1$.

We show the face SR results on several examples. In Fig. 4, we can see that the results of our method show fewer artifacts and are more smoothness than that of other algorithms involved in the comparison. Especially, our method is stable on different LR image sizes. It is due to that our method manifold alignment could reduce the conflict between the manifold assumption and the case of the original couple of LRM and HRM which is caused by the nonisometric mapping from HR to LR.

In Fig. 5, we show the SR results of different methods on LR face images of size 18×18 pixels. It is worth noting that our method outperforms other methods in areas close to eyes and mouths. The results shown in the first row of the figure is fairly interesting. While other comparative learning-based methods synthesize the lips as teeth, the result of our method is more reasonable and close to real lips.

Furthermore, we compare the mean root mean square (RMS) errors of our results with other methods. From Table II, we can see that our method achieves lower RMS errors and gets the better or comparable quantitative results. It means that our method could estimate the missing high-frequency information more correctly.

TABLE II
RMS ERRORS OF DIFFERENT METHODS

LR/HR size	Bi-cubic	Freeman et al., [4]	Chang et al., [3]	Yang et al., [14]	Ours
12/72	11.13	10.49	9.48	**9.14**	9.15
18/72	7.85	8.10	7.39	7.21	**7.06**
24/72	6.32	6.61	6.06	5.91	**5.80**

V. CONCLUSIONS

In this paper, we revisit the manifold assumption which is the basis of some learning-based SR algorithms. From the quantitative analysis, we find that the manifold assumption does not hold well on the original LRM and HRM. To overcome the limitations of previous methods, we propose a novel method to align original LRM and HRM into a CM. After the alignment, the local geometry of CM is more consistent with HRM than LRM is. Then the SR reconstruction based on neighbor embedding takes advantage of the manifold alignment and achieves better results. The experimental results reveal the effectiveness of proposed method. A problem of our method is that the proposed linear method is hard to process more complex problems, such as natural image SR. Using kernel technique to extend our method for solving the nonlinear problems will be pursued in our future work. Besides, we will try to use tensor-based representation and prediction method [12] to improve our method.

REFERENCES

[1] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *Proc. IEEE Int. Conf. Image Processing*, 1996, pp. 707–710.
[2] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 1167–1183, 2002.
[3] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004, pp. 275–282.
[4] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, pp. 25–47, 2000.
[5] W. Fan and D. Y. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–7.
[6] J. Ham, D. Lee, and L. Saul, "Semisupervised alignment of manifolds," in *Proc. Tenth Int. Workshop on Artificial Intelligence and Statistics*, 2005, pp. 120–127.
[7] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, pp. 324–335, 1993.
[8] C. Liu, H. Shum, and C. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 192–198.
[9] P. Philips, H. Moon, P. Pauss, and S. Rivzvi, "The feret evaluation methodology for face-recognition algorithms," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1997, pp. 137–143.
[10] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
[11] J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. 729–736.
[12] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. J. Maybank, "Bayesian tensor approach for 3-D face modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, pp. 1397–1410, 2008.
[13] Y. W. Tai, W. S. Tong, and C. K. Tang, "Perceptually-inspired and edge-directed color image super-resolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006, pp. 1948–1955.
[14] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.