This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

SLR: Semi-coupled locality constrained representation for very low resolution face recognition and super resolution

Tao Lu, Xitong Chen, Yanduo Zhang, Chen Chen and Zixiang Xiong

***Abstract*—Although face recognition algorithms have been greatly successful recently, in real applications of very low resolution (VLR) images, both super resolution (SR) and recognition tasks are more challenging than those in high-resolution (HR) images. Given the rare discriminative information in VLR images, the one-to-many mapping relationship between HR and VLR images degrades the SR and recognition performances. In this paper, we propose a novel semi-coupled dictionary learning scheme to promote discriminative and representative abilities for face recognition and SR simultaneously by relaxing coupled dictionary learning. Specifically, we use semi-couple locality-constrained representation to enhance the consistency between VLR and HR local manifold geometries, thereby overcoming the negative effects of one-to-many mapping. Given the learned task-oriented mapping function, we feed these discriminative features into a collaborative representation-based classifier to output their labels, and combine a locality-induced approach to hallucinate the HR images. Extensive experimental results demonstrate that the proposed approach outperforms a number of state-of-the-art face recognition and SR algorithms.**

***Index Terms*—Very Low-resolution, semi-coupled locality-constrained representation, face recognition, face hallucination.**

I. Introduction

Recently, face recognition algorithms have made significant progresses in many real-world applications, for instance surveillance video, authentication and entertainment. However, in real scenarios, the resolution of facial images is often in very low because of the far distance between cameras and targets, thereby these thumb-size very low resolution (VLR) facial images brings huge challenges to existing face recognition algorithms. In addition, variations of pose, illumination and expression make the recognition even more difficult than high-resolution (HR) images [1], [2], [3], [4]. In this paper, we define that the resolution of VLR face image is no more than 10×10 pixels.

Usually, there are three types of current VLR face recognition algorithms. The first type is downsample-based approaches, in which HR gallery images are down-sampled, thereby matching input low-resolution (LR) testing samples. The second type is super-resolution (SR) based approaches, in which the LR probe is upscaled, thereby matching the HR gallery. The third type is simultaneous super-resolution and recognition (SRR)-based approaches, which completes SR and recognition tasks at the same time. Down-sampling HR gallery image into LR version easily solves the resolution mismatching problem. However, degradation process of down-sampling degrades the available information. Then, SR algorithms are generally applied in magnifying the resolution of LR probe images to overcome the HR and VLR resolution gap [5].

Generally, we can easily use SR algorithms to enlarge the VLR features into the size of HR images for VLR face recognition. In the past few decades, various learning-based face hallucinations are proposed to render pleasure visual results, thereby overcoming the resolution gap of HR gallery images and VLR probe ones. Two types of face SR algorithms are widely used, namely, vision-based and feature-based.

Baker [6] *et al.* first proposed the concept of "face hallucination", which refers a subject-specific SR algorithm. They used resolution pyramid to match the different resolution small patches. After this remarkable work, Wang [7] *et al.* used " Eigenface" to transform LR image features into HR space for reconstruction. Chang [8] *et al.* first introduced neighbor embedding which also named as locally linear embedding (LLE) into face SR algorithms. Then, Ma [9] *et al.* assumed same position-patch has similar content similarity and used least squares representation (LSR) to represent patch prior. To overcome the over-fitting, Jung [10] *et al.* used sparse regularization term to constrain the representation weights and achieved good subjective results (SRSR). Jiang [11], [12] *et al.* used locality-constrained representation (LCR) to regularize representation weights by exploring the manifold structure. Shi [13], [14] *et al.* developed a framework of face hallucination using global image-level consistency, local patch sparsity, pixel-level correlation and kernel prior. For the noisy input images, low-rank representation [15], [16] were used in promoting the robustness of representation coefficients. Deep collaborative representation [17] and deep linear mappings learning [18] extended the representation ability for training samples. In recent years, Dong [19] *et al.* firstly proposed three-layered convolutional neural networks for SR (SRCNN). Region-based convolutional neural networks [20] leveraged face super-resolution performance by accurate priors. Shi [21] *et al.* discussed the roles of regularization models in HR feature space. Noise-robust hallucination [22] and nonlocal structure prior [23] were discussed for better performance.

T. Lu, X. Chen, and Y. Zhang are with Hubei Key Laboratory of Intelligent Robot, School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, 430070, China. (e-mail:lutxyl@gmail.com).

C. Chen is with the Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223 USA (e-mail: chenchen870713@gmail.com).

Z. Xiong is with the Department of Electrical and Computer Systems Engineering, Texas A&M University, College Station, TX 77843 USA(e-mail:zx@ece.tamu.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

The above vision-based SR algorithms have superior subjective/objective performances by minimization of reconstruction errors. Nevertheless, they can not fully consider the discriminative ability of features, thereby degrading their recognition performance capability.

For improved recognition rates, feature-based SR approaches are proposed, thereby obtaining resolution-robust features that can ameliorate the resolution gap between LR and HR features. Li [24] *et al.* aligned different resolution features into a consistent manifold feature space for enhancing feature representation ability. Jiang [25] *et al.* used coupled discriminant multi-manifold to analyses (CDMMA) the VLR face features for enhancing the difference of different manifolds. This approach learns two discriminant projective matrices to transfer VLR and HR features into a common feature space. A resolution-invariant image representation scheme had been proposed to match the LR image patch to an HR one [26]. Wang [27] *et al.* reviewed different deep structures of deep networks for LR recognition. Maeng [28] proposed a cross-mode matching method to perform VLR image recognition. Shi [29] *et al.* learned a latent subspace for aligning HR and LR features for improving recognition performance. Ma [30], [31] *et al.* used robust matching to improve registration. Fast matching for non-rigid image feature with probabilistic inference was proposed for UAV [32]. Low-rank supported extreme learning machine yielded robust performance against noise and outliers [33]. Although these feature-robust based SR methods indicate satisfactory performance in recognition, they do not consider image reconstruction because they consider HR features instead of HR images.

To rectify the above shortcomings, several algorithms focus on the recognition and hallucination tasks at the same time. They aim to improve the recognition rate and the visual performance simultaneously. Hennings-Yeomans [34] *et al.* combined SR and recognition tasks in one algorithm (S^2R^2) for promoting VLR face image recognition performance. Biswas [35], [36] *et al.* introduced multi-dimensional scaling (MDS) to preserving the distance in HR feature space. For far distance observed small image, Yang [37] *et al.* used sparse regularization two times for both VLR face recognition and hallucination (FRH). Huang [38] *et al.* modeled the mapping relationship of VLR and HR features by canonical correlation analysis (CCA). Jian [39] *et al.* used LR and HR singular values to build a feature relationship.

Most simultaneous SR and recognition algorithms assume that VLR and HR features have similar manifold geometry structure. However, in VLR scenario, using this assumption is difficult. As shown in Fig. 1, we list the structure preservation rate [40], [41] of three types of resolution face images, from which, the structure preservation rate decreases from 95% to 73% for 2× downsampling to 8× downsampling (where $K = 4$). Structure preservation rate measures the neighborhood relationship between HR and LR images, i.e., for images that are neighbors in HR image space, their LR versions are also neighbors as well. One-to-one mapping indicates a 100% structure preservation rate. This phenomenon illustrates that LR and HR structure mismatching degrades the VLR recognition performance. Thus, VLR has weak ability to distinguish

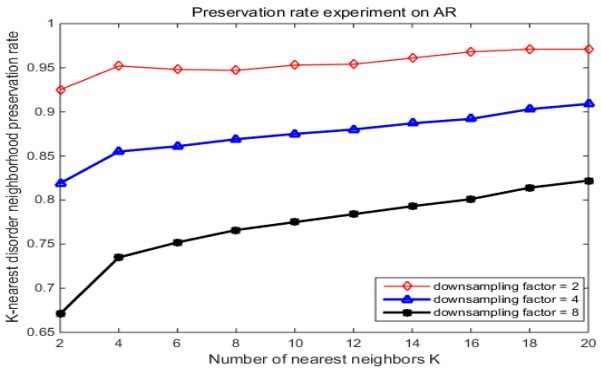

Fig. 1. Discriminative feature preservation rate indicates degree of LR and HR structural matching. K is the number of subclass in gallery for AR database.

the identity of different individuals by limited features. The credible discriminative information of VLR images are lost during degradation process. Many works [42], [43], [44] have pointed out that image resolution is not the vital factor in face recognition systems. By contrast, the discriminative features learned from dictionary is the dominant factor. In fact, both upsampling- and downsampling-based algorithms can not perfectly extract the discriminative features. Thus, the one-to-many complex mapping function between HR and VLR features degrade both the reconstructive and discriminative performances of images. Most VLR face recognition approaches obey the manifold consistency assumption. They simplify the complex relationships between HR and VLR into linear models. However, these complex mapping functions between VLR and HR images are very hard to represent in the full coupled learning scheme. Recently, semi-coupled dictionary learning methods [45], [46] have been developed for revealing the complex relationship of VLR-to-HR features.

Motivated by the above studies [45], [46], we use a semi-coupled locality-constrained representation (SLR) to revise the one-to-many mapping functions for improving representation ability. First, we replace sparse regularization term by locality-constrained representation (LCR) term which has powerful discriminative feature representation ability. Then, we propose a novel semi-couple framework to learn a dictionary pair to transform the VLR features into their HR version. When the mapping functions are ready, the transformed VLR features are robust to varying resolution. Finally, we use a collaborative representation-based classifier to perform the recognition task. Extensive experiments show that SLR outperform many state-of-the-art VLR face recognition methods.

We have extended our preliminary work [47] to simultaneous recognition and SR. First, a VLR face image SR algorithm is added into the semi-coupled learning scheme to further prove the proposed method. Second, we provide additional experimental result details about the proposed semi-coupled learning scheme. Finally, we introduce methods of parameter selection to guide fine-tuning by maximum a posteriori probability estimation. We summarize the contributions of this paper as follows:

1) We proposed a semi-coupled dictionary learning method

for VLR face image feature representation and mapping. The learned LR features are transformed into HR space for simultaneously recognition and hallucination.

2) We comprehensively analyzed the role of locality-constrained representation in recognition and SR tasks including selection of parameters and optimization of its semi-coupled version.

The structure of this paper is as follows: section I introduces the VLR simultaneous recognition and SR problem. Section II reviews relevant representation schemes, such as sparse representation, locality-constrained representation, and coupled learning scheme. Section III provides the objective function and its optimization of semi-coupled locality-constrained representation frameworks. Section IV conducts comprehensive experiments to investigate the proposed approach. The last section concludes.

II. Related work

A. Sparse Representation

Given an over-completed dictionary codebook $D = [\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_M] \in \Re^{d\times M}$, M indicates dictionary atoms amount, for an input d-dimension data vector $\boldsymbol{x} \in \Re^{d\times 1}$, the sparse representation weights can be optimized by objective function:

$$\arg\min_{\boldsymbol{\alpha}} = \{||\boldsymbol{x} - D\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_0\}, \quad (1)$$

here $\boldsymbol{\alpha}$ is the M-dimension coding vector, $||\cdot||_0$ represents ℓ_0-norm which counts the amount of non-zero atoms, λ is a controlling parameter that contributes to the regularization. Given the NP-hardness of ℓ_0-norm, ℓ_1-norm ($||\boldsymbol{\alpha}||_1 = \sum_i^M |\alpha_i|$) is typically used to substitute ℓ_0-norm in optimization. Most of existing sparse learning approaches are based on ℓ_1-norm regularization because of its sparsity-inducing property, convenient convexity, strong theoretical basis, and significant success in many scenarios, i.e., multi-modal sparse coding was successfully applied to web ranking [48].

B. Locality-constrained Representation

LCR considers the manifold structure of input signal and dictionary atoms, thereby resulting in improved recognition performance. The locality constraint is more significant than sparsity in revealing the true geometry of a nonlinear manifold [23]. The objective function incorporates a locality constraint as follows:

$$\arg\min_{\boldsymbol{\alpha}}\{\|\boldsymbol{x} - D\boldsymbol{\alpha}\|_2^2 + \lambda\,\|\boldsymbol{l}\circ\boldsymbol{\alpha}\|_2^2\}, s.t. \mathbf{1}^T\boldsymbol{\alpha} = 1, \quad (2)$$

here $\circ$ denotes an element-wise vector product, while $\boldsymbol{l} = [l_1, ..., l_M]^T$ is an M-dimensional vector representing the Euclid distance of query patch and every dictionary atoms. Specifically,

$$l_i = \exp(\frac{dist(\boldsymbol{x}, D)}{\sigma}), \quad (3)$$

here $dist(\boldsymbol{x}, D) = [dist(\boldsymbol{x}, d_1), ..., dist(\boldsymbol{x}, d_M)]^T$, and $dist(\boldsymbol{x}, d_j)$ is the Euclidean distance between x and d_j, σ adjusts the decay speed of the locality adaptor; $\mathbf{1}^T$ is an all-one column vector, and constraint $\mathbf{1}^T\boldsymbol{\alpha} = 1$ guarantees shift-invariance. We normalize l_i between $(0, 1]$. Furthermore, λ denotes the regularization controlling parameter. The regularized least squares method is selected to analytically derive $\boldsymbol{\alpha}$. Compared with the SR, the LCR improved the reconstruction, and local smooth sparsity. Moreover, the analytical solution of LCR are widely used in various computer vision applications.

C. Coupled Sparse Representation

In transfer learning scenario, i.e., face SR, LR recognition, coupled sparse representation (CSR) is widely used to ameliorate the information gap between different resolution data. For D_l and D_h represent LR and HR dictionaries, $\boldsymbol{x}_l$ and $\boldsymbol{x}_h$ represent training data vector. The coupled sparse representation seeks a resolution invariance weights by the following objective function,

$$\arg\min_{\boldsymbol{\alpha}} = \{||\boldsymbol{x}_l - D_l\boldsymbol{\alpha}||_2^2 + ||\boldsymbol{x}_h - D_h\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1\}, \quad (4)$$

where $\boldsymbol{\alpha}$ is a coupled sparse weighting vector across different resolution domains. Essentially, CSR follows the manifold consistency assumption, i.e., LR and HR data share an isometric local manifold structure (isometric representation weights). Formula (4) can be solved as a typical SR optimization after coupling the input vectors $\boldsymbol{x} = [\boldsymbol{x}_l, \boldsymbol{x}_h]^T$, and dictionaries $D = [D_l, D_h]^T$ together. In many applications, CSR obtains satisfactory results from image reconstruction to recognition tasks [46].

III. The proposed method

Although coupled-learning scheme effectively models one-to-one mapping and are greatly successful in many applications, it restricts both representation and reconstruction abilities in different domains. To simultaneously boost representative and discriminative capabilities, semi-coupled learning approaches [37], [45], [46] are proposed for overcoming the limitation of coupled-learning. Inspired by these studies, we use semi-coupled locality-constrained representation to enhance the discriminative and representative abilities in both the VLR and HR domains.

The outline of SLR is shown in Fig. 2. Semi-coupled learning scheme includes two parts, namely, the training and testing phases. The aim of training phase is to learn two (VLR and HR) dictionaries and the mapping matrix W. Given these clues, an engine integrates simultaneous face recognition and SR into two algorithms. Thus, the final outputs indicate two different channels. One is HR image, and the other is the identity information of the query image. It is believed that VLR images indicate lower recognition performance because they contain rare discriminative information. Given the degradation, the LR versions of two totally different HR images may look similar, thereby leading to one-to-many mapping in VLR scenarios.

Based on manifold learning, HR and LR images share isometric geometric structures as per the manifold consistency hypothesis. Traditional sparse representation approaches [10], [49], [50] used the coupled learning model to support the hypothesis. However, in the VLR scenario, the hypothesis is difficult to support; the semi-coupled learning scheme relaxes the strict and strong constraints on coupled learning

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

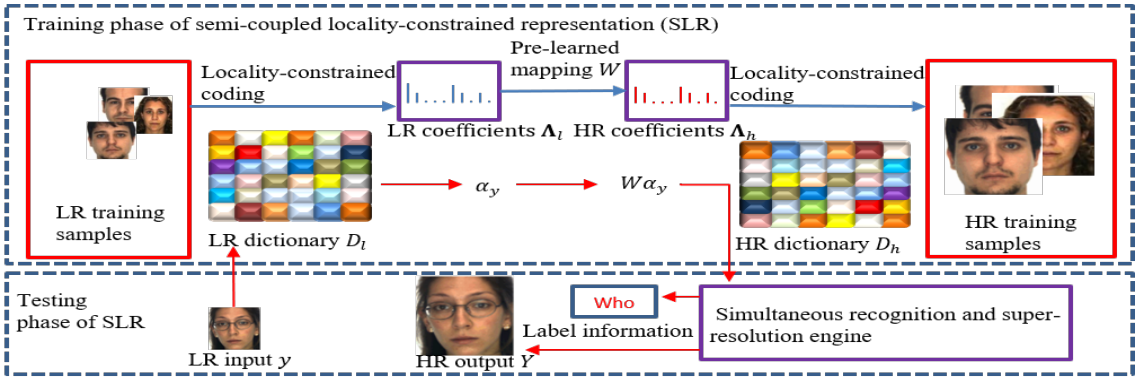

Fig. 2. Flowchart of the proposed semi-coupled locality-constrained representation method for face simultaneous recognition and SR. The blue arrows indicate training process and the red arrows represent testing steps.

for flexible and accurate representation ability. Furthermore, category information in dictionaries promote discriminative and representative abilities by transforming mapping function W by correcting the degraded manifold structure as shown in Fig. 2.

A. Semi-coupled Locality-constrained Representation

We use LCR to represent the HR and VLR samples pairs in the same time, thereby enhancing the representation capability of discriminative features.

Let $X_l = [\boldsymbol{x}_l^1, ..., \boldsymbol{x}_l^n] \in \mathbb{R}^{d\times n}$ represents the VLR d-dimension image dataset, n is the number of VLR images; $D_l = [\boldsymbol{d}_l^1, ..., \boldsymbol{d}_l^M] \in \mathbb{R}^{d\times M}$ is the VLR dictionary; M is the amount of dictionary atoms; $\Lambda_l = [\boldsymbol{\alpha}_l^1, ..., \boldsymbol{\alpha}_l^n] \in \mathbb{R}^{M\times n}$ is the VLR coefficient matrix. Correspondingly, $X_h = [\boldsymbol{x}_l^1, ..., \boldsymbol{x}_l^n] \in \mathbb{R}^{t\times n}$, $D_h = [\boldsymbol{d}_h^1, ..., \boldsymbol{d}_h^M] \in \mathbb{R}^{t\times M}$ and $\Lambda_h = [\boldsymbol{\alpha}_h^1, ..., \boldsymbol{\alpha}_h^n] \in \mathbb{R}^{M\times n}$ represent the HR dataset, dictionary, and LCR representation coefficients matrices. Furthermore, $t = d \times s^2$ and s is the amplification factor. To obtain the LR and HR representation coefficients and their relationship regression function, different from LCR, we design four regularization terms named as VLR dictionary term E_L, HR dictionary term E_H, mapping term E_M and as mapping constrain term E_R. Thus, total objective function is represented as following:

$$\begin{aligned}&\arg\min_{\Lambda_l,\Lambda_h f(\cdot)}\{E_L\left(X_l, D_l\right) + E_H\left(X_h, D_h\right)\\&+E_M\left(\Lambda_h, f\left(\Lambda_l\right)\right) + E_R\left(f\left(\cdot\right)\right)\}\end{aligned}, \tag{5}$$

where E_L and E_H indicate the image reconstruction error; E_M represents the mapping error between the VLR and HR representation coefficients. E_R represents the regularization of the mapping matrix, and $f(\cdot)$ is the mapping function between the VLR and HR features matrices.

To simplify the complex mapping function, we use $f(\cdot)$ as a linear mapping and rewrite it as W. Then, objective function (5) can be represented by:

$$\begin{aligned}&\arg\min_{D_h,D_l,\Lambda_h,\Lambda_l,W}\{(\|X_l - D_l\Lambda_l\|_F^2 + \|X_h - D_h\Lambda_h\|_F^2\\&+\lambda_1\sum_{i=1}^{n}\left\|l_l^i \circ \alpha_l^i\right\|_2^2 + \lambda_2\sum_{i=1}^{n}\left\|l_h^i \circ \alpha_h^i\right\|_2^2 + \lambda_3\|\Lambda_h - W\Lambda_l\|_F^2\\&+\lambda_4\|W\|_F^2)\}\end{aligned}, \tag{6}$$

where α_l^i and α_h^i are the $i-$th VLR and HR representation coefficients vectors; $\circ$ represents element-wise vector product, l_l^i and l_h^i are M-dimensional weighted vectors that represents the Euclidean metric of input patch and their representation dictionary atoms in different resolution domains respectively. Four λs are the balance parameters to adjust the contributions ratios of the four regularization terms. In the proposed semi-coupled learning scheme, we simultaneously learn representation dictionaries and their coefficients mapping functions. First, VLR images will represented by VLR dictionary. Then with the learned mapping function, the LR coefficients will be projected into their HR coefficients as features. Thus these learned features can been known as having resolution-robust and discriminative abilities. We use alternating iteration algorithm to solve the above objective function.

B. Optimization of SLR scheme

Given two training datasets, the purpose of SLR is to learn the semi-coupled dictionary pair D_l , D_h ,and the mapping relationship W. First, we use VLR and HR data matrices as initial D_L , D_h. W is initialized as an identity matrix, and λ_1, λ_2 are initialized at 0. We iteratively update Λ_h and Λ_l, D_l and D_h, W as following steps.

Updating Λ_l and Λ_h: When HR X_h and LR X_l, the dictionary pair D_l, D_h and λ are ready. Then, we solve the LCR coefficients Λ_l, Λ_h individually as follows:

$$\begin{aligned}&\arg\min_{\boldsymbol{\alpha}_l^i}\sum_{i=1}^{n}\{\left\|\boldsymbol{x}_l^i - D_l\boldsymbol{\alpha}_l^i\right\|_2^2 + \lambda_1\left\|\boldsymbol{l}_l^i \circ \boldsymbol{\alpha}_l^i\right\|_2^2 +\\&\lambda_3\left\|\boldsymbol{\alpha}_h^i - W\boldsymbol{\alpha}_l^i\right\|_2^2\},\end{aligned} \tag{7}$$

$$\begin{aligned}&\arg\min_{\boldsymbol{\alpha}_h^i}\sum_{i=1}^{n}\{\left\|\boldsymbol{x}_h^i - D_h\boldsymbol{\alpha}_h^i\right\|_2^2 + \lambda_2\left\|\boldsymbol{l}_h^i \circ \boldsymbol{\alpha}_h^i\right\|_2^2 +\\&\lambda_3\left\|\boldsymbol{\alpha}_h^i - W\boldsymbol{\alpha}_l^i\right\|_2^2\}.\end{aligned} \tag{8}$$

To resolve the above two formulas, we rewrite locality regularization terms into matrix forms, $L_l^i = diag(\boldsymbol{l}_l^i)$ indicates the Euclidean distance of input $\boldsymbol{x}_l^i$ and dictionary atom $\boldsymbol{d}_l^j$. We use the solution of LR representation matrix Λ_l as an example, Formula (7) can be replaced as:

$$\begin{aligned}&\arg\min_{\boldsymbol{\alpha}_l^i}\{\sum_{i=1}^{n}\left\|\boldsymbol{x}_l^i - D_l\Lambda_l\right\|_2^2 + \lambda_1\left\|L_l^i\boldsymbol{\alpha}_l^i\right\|_2^2 +\\&\lambda_3\left\|\boldsymbol{\alpha}_h^i - W\boldsymbol{\alpha}_l^i\right\|_2^2\},\end{aligned} \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

We combine the first and third terms of the equation. The function can thus be rewritten as:

$$\arg\min_{\boldsymbol{\alpha}_l^i}\{||\begin{bmatrix} \boldsymbol{x}_l^i \\ \sqrt{\lambda_3}\boldsymbol{\alpha}_h^i \end{bmatrix} - \begin{bmatrix} D_l \\ \sqrt{\lambda_3}W \end{bmatrix}\boldsymbol{\alpha}_l^i||_2^2 + \lambda_1||L_l^i\boldsymbol{\alpha}_l^i||_2^2\}. \tag{10}$$

In this formula, only $\boldsymbol{\alpha}_l^i$ is unknown while the other variables are known in advance. We let $H = \begin{bmatrix} D_l \\ \sqrt{\lambda_3}W \end{bmatrix}$, and $y = \begin{bmatrix} x_l^i \\ \sqrt{\lambda_3}\alpha_h^i \end{bmatrix}$. Thus, this function can be resolved with a regularized least square with the analytic solution:

$$(\boldsymbol{\alpha}_l^i)^* = (H^TH + \lambda_1 L_l^i)^{-1}H^T\boldsymbol{y}. \tag{11}$$

Then, we can solve Λ_h in series in the same manner.

Updating D_l and D_h: When LR and HR representation coefficients matrices are given, the dictionaries D_h and D_l can be updated as follows:

$$\arg\min_{D_l}\{||X_l - D_l\Lambda_l||_F^2 + \lambda_1\sum_{i=1}^{n}\left\|l_l^i \circ \alpha_l^i\right\|_2^2\}, \tag{12}$$

$$\arg\min_{D_h}\{\|X_h - D_h\Lambda_h\|_F^2 + \lambda_2\sum_{i=1}^{n}\left\|l_h^i \circ \alpha_h^i\right\|_2^2\}. \tag{13}$$

Similar to a previous work [51], we excluded the locality constraint term to simplify the computation. We use gradient descent algorithm to solve the above problem. The updating scheme for D_l is $D_l^{j+1} = D_l^j - \beta\nabla D_l^j$ and $\nabla D_l^j = -2(X_l - D_l^j\Lambda_l)\Lambda_l^T$. Here, β is a step size length that controls the learning rate. Thus, D_h can be derived in the same manner.

Updating W: As soon as coefficients matrices Λ_l and Λ_h, and dictionaries D_l and D_h are ready, mapping function W can be updated as follow:

$$\min_W\{\lambda_3\|\Lambda_h - W\Lambda_l\|_F^2 + \lambda_4\|W\|_F^2\}, \tag{14}$$

this objective function is an unconstrained quadratic problem. Let $G(W)$ denote the above objective function. Then, let

$$\frac{\partial G(W)}{W} = (-2\Lambda_h\Lambda_l^T + 2W\Lambda_l\Lambda_l^T + 2\tfrac{\lambda_4}{\lambda_3}W) = 0. \tag{15}$$

The function above can be rewritten as follow, where W is initialized as an identity matrix:

$$W = \Lambda_h\Lambda_l^T(\Lambda_l\Lambda_l^T + \tfrac{\lambda_4}{\lambda_3}I)^{-1}, \tag{16}$$

where I is the identity matrix.

We iteratively perform the above steps for few iterations, D_l and D_h, W can reach a balance to represent VLR images and transform them into HR feature space. We directly use these learned dictionary pair and mapping matrix in the following classification task. The summary of semi-coupled learning scheme is in Algorithm.1.

Algorithm 1 Dictionary learning by semi-coupled scheme.

Input: Validation of datasets X_h, X_l. Initial training dictionary D_h, D_l,

1: Initial mapping of coefficients W.
2: Fix other variables, update $\boldsymbol{\alpha}_h$ and $\boldsymbol{\alpha}_l$ by Eq.(7) and Eq.(8).
3: Update D_h and D_l by Eq.(12-13) with fixed other variables
4: Update W by Eq.(16).
5: Repeat the above steps with set iteration number.

Output: Semi-coupled dictionaries D_h and D_l and mapping coefficients W.

C. Classification

In the classification step, the codebook D contains n training images of k object classes: $D = [D_1, D_2, ...D_k]$, $n = \sum_{i=1}^{k} n_i$. With the i-th face class $D_i = [\boldsymbol{d}_{i,1}, \boldsymbol{d}_{i,2}...\boldsymbol{d}_{i,n_i}] \in \Re^{t\times n_i}$, n_i is the number of the i-th class. We directly take HR dictionary D_h as the gallery database. For the query LR input y, we first obtain its LR representation coefficient $\boldsymbol{\alpha}_y$. Then the VLR features are transformed into HR domain by $\boldsymbol{\alpha}_h = W\boldsymbol{\alpha}_y$.

For the i-th class, we define δ_i as a selection function which choose the coefficients according to the same class. For input query VLR image y, we classify it by assigning it to the object class that minimizes its construction errors as:

$$\arg\min_i r_i(\boldsymbol{y}) = \min_i\{\|\boldsymbol{y} - D\delta_i(W\boldsymbol{\alpha_y})\|_2^2 / ||W\boldsymbol{\alpha}_y||_2^2\}. \tag{17}$$

This objective function has analytical solution by using the SPAMS toolbox [52].

D. Locality-induced based face hallucination

In the face hallucination scenario, traditional face hallucination algorithms directly project the LR representation weights to the HR dictionary space for reconstruction. However, the "one-to-many" approach and related degradation reduces the reconstruction ability of LR dictionary. Especially in a VLR scenario, the structure mismatching of representation weights between LR and HR manifolds reduces both recognition and reconstruction efficiency. To use the locality constraint in the super-resolved HR target image, we rewrite (2) into

$$\begin{aligned} &\arg\min_{\alpha}\|\boldsymbol{y} - D\boldsymbol{\alpha}\|_2^2 + \lambda\sum_{i=1}^{n}\|l_i \circ \alpha_i\|_2^2, \\ &s.t.\sum_{i=1}^{n}\alpha_i = 1, and\ \alpha_k = 0\ if\ k \notin C_k(y), \end{aligned} \tag{18}$$

where $C_k(y)$ is the indices of the K nearest neighbors of $\boldsymbol{y}$ in the VLR dictionary. Given the added constraint, the value of α_i shrinks to zero if d_i is not the K nearest neighbors of $\boldsymbol{y}$, Thus, we only need to determine the K nearest neighbors and use them to reconstruct $\boldsymbol{y}$. Here, we use $\varepsilon - ball$ method to find K neighbors: If $||\boldsymbol{d}_i - \boldsymbol{y}||_2^2 < \varepsilon$, then add index i into $C_k(y)$, where $\varepsilon > 0$ is a constant. Formula (18) can be seen as a regularized least square problem, which has an analytical solution [12], [53].

Thus, we summarize the above classification process in Algorithm.2.

IV. EXPERIMENTS

A. Database configuration and parameter settings

We perform extensive experiments on AR [54] and CMU PIE [55] Face databases to investigate the performance of the SLR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

Algorithm 2 Simultaneous face recognition and SR engine via SLR.

Input: Semi-coupled dictionaries D_h, D_l, mapping function W, and testing sample $\boldsymbol{y}$.

VLR face recognition

1: Calculate the testing LR LCR coefficients $\boldsymbol{\alpha}_l$.
2: Transform $\boldsymbol{\alpha}_l$ to $\boldsymbol{\alpha}_h$ by mapping function W.
3: Classify with Eq (17).

VLR face hallucination

4: Calculate K locality-induced dictionary for input LR $\boldsymbol{y}$.
5: Use (18) to calculate the LR LCR coefficient $\boldsymbol{\alpha}_l$.
6: Transform $\boldsymbol{\alpha}_l$ into $\boldsymbol{\alpha}_h$ with the mapping function W_K.
7: Calculate HR image by $Y = D_h W_K \boldsymbol{\alpha}_l$.

Output: Class label i for input $\boldsymbol{y}$, and HR image Y.

Fig. 3. Visual samples from the AR database. First row: HR samples; Second row: LR samples.

The AR database has 100 subjects and total 2,600 face images that include varied expressions and illumination. We drop the images of subjects wearing glass and scarf, and select 1,400 samples include expression and illumination changes. The entire database is partitioned into three parts. The first part (five images for each person, totally 500 samples) is used as LCR initial dictionaries D_h and D_l. The second part (five other 5 images for each person, totally 500 samples) is used as the training set (training datasets X_h and X_l) for the mapping coefficients. We use the remainders (dataset Y_l contains 400 samples) to testify the face recognition performance. As we know, all test images are different from the training database. The original AR database are color images, all images are 768×576 pixels. We manually align the extracted faces images from the AR database by the positions of the two eyes. Then we transform the color images to grayscale images and resize them to 32×24 pixels as HR face images by bicubic interpolation. We downsample the HR by a bicubic interpolation at 1/4 multiple. The size of VLR image is 8×6 pixels. Some representative samples from AR database are shown in Fig. 3.

The CMU PIE database have 68 subjects and total 40,000 facial images, including poses, illumination, and expressions changes. We randomly divide the whole database into three parts. The first part (5 images for each person, totally 1360 samples) is used as LCR initial dictionaries D_h and D_l. The second part (20 other images for each person) is used as the training samples (training datasets X_h and X_l, each dataset has 340 samples) for the mapping coefficients. We select the third part (dataset Y_l, with 28 images per person, with certain extremely dark images are removed, totally 1969 samples) to evaluate face recognition performance. The original CMU

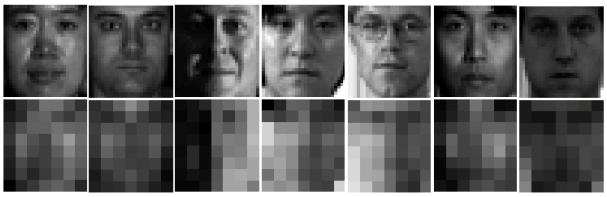

Fig. 4. Visual samples of the CMU PIE database. First row: HR samples; Second row: LR samples.

PIE database are colorful human faces images, all images are 640×486 pixels. we directly convert the color images to grayscale images and then resize the grayscale images to 32×28 pixels as HR face images by bicubic interpolation. We downsample the HR image by bicubic interpolation with amplification factor of 1/4. Then the size of VLR image is 8×7 pixels. We show the visual HR and LR samples from CMU PIE database in Fig. 4.

First, we initialize D_h and D_l as HR face image data matrices and VLR ones respectively. The number of iterations is 20. In the AR database we fine-tune balance parameters λ_1=0.02 and λ_2=0.06, λ_3=1 and λ_4=0.01. In the CMU PIE database we use balance parameters $\lambda_1=2 \times 10^{-5}$ and $\lambda_2=8 \times 10^{-3}$, $\lambda_3=6.4 \times 10^{-5}$ and $\lambda_4=10^{-3}$. For the VLR image recognition task, dataset X_h is selected as the HR gallery set, and Y_l is used as the VLR probe set. There are no overlapped images between the training data and testing data. Here, for each image in gallery dataset, we use HR dictionary D_h to extract features and for LR probe dataset, we use LR dictionary D_l to extract LR features, with the learned mapping function, we can transform the LR features into HR ones. For a fair comparison, some state-of-the-art recognition and SR algorithms are chosen as the benchmarks for VLR face images. As far as we know, the best performances have been reported for the following: FRH [37] presents for VLR face recognition and hallucination; CDMMA [25] for feature-based face recognition; VDSR [56] for vision-based face hallucination with deep learning; and CNE [41] for representation-based face hallucination. All of the best performance parameters have been considered in this work. Furthermore, all of the face recognition and SR experiments are conducted over the same database settings.

B. Effect of semi-coupled learning scheme (mapping function W)

On the basis of transfer learning, the correlation of different knowledge domains determines the learning performance [57]. As typical applications of transfer learning, SR- and representation-based recognition tasks rely on the accuracy of representation coefficients. Here we use the accuracy index of representation coefficients the same way as in previous work [16] for quantitative measurement. Suppose, the LR input $\boldsymbol{y}^l$ and its original HR image $\boldsymbol{y}^o$. Then, their coefficients are $\boldsymbol{\alpha}^l$ and $\boldsymbol{\alpha}^o$. In reality, the HR coefficients of $\boldsymbol{\alpha}^o$ are not available, but we can use the original image $\boldsymbol{y}^o$ to test the coefficient accuracy. We use the correlation value of $(\boldsymbol{\alpha}^0)^T \boldsymbol{\alpha}^l$ as accuracy metric. The proof details can be read in [16]. As

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

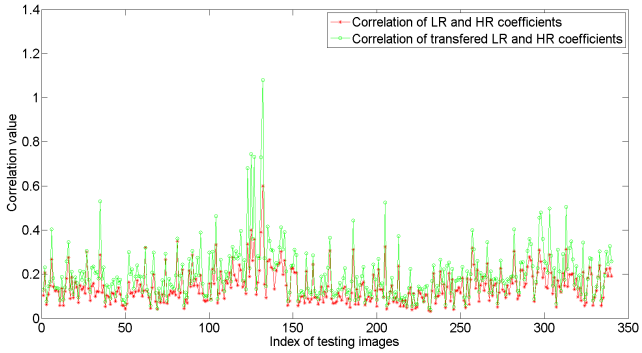

Fig. 5. Correlation value denotes the consistency of representation coefficients. Larger value indicate better accuracy for coefficients. X-axis represents the index of 360 testing images whereas Y-axis represents the correlation values. Experiments are conducted using the AR face database, and the configuration is the same as in the next experiment.

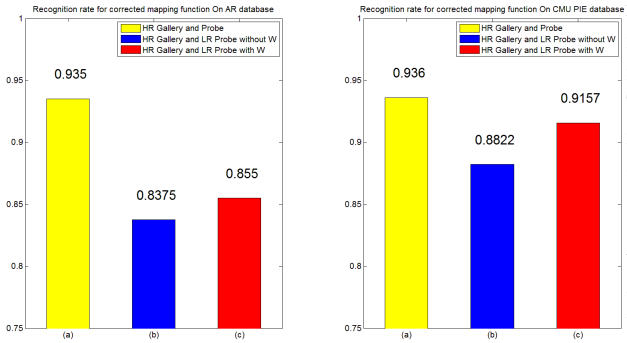

Fig. 6. Recognition rate with or without the mapping function W. Left: AR database. Right: CMU PIE database.

shown by the simulation results in Fig. 5, given the transform function W, the accuracy of representation coefficients is better than original LR and HR coefficients using the semi-coupled learning scheme.

Another experiment is designed to further verify the effectiveness of the mapping function, one with and the other without the W matrix. When the best reconstructive and discriminative dictionary pair is learned, we use the gallery and probe sets all in HR size as the upper bound. The performance of the proposed algorithm with different setting is shown in Fig. 6. The learned semi-coupled mapping matrix enhances the average recognition rate by 1.75% for the AR database and 3.35% for the CMU PIE database. This results confirm that the proposed semi-coupled dictionary learning scheme is important in FR task. Mapping functions can ameliorate one-to-many relationships by local manifold regularization. The resolution gap between HR and LR is reduces by 5% - 10% in terms of recognition performance, and the transforming function W improves the performance by revising the manifold structure.

C. Effect of LCR

Unlike sparse representation regularization, LCR shows some attractive advantages, for instance enhanced reconstruction ability, local smooth sparsity and having analytical solution [11]. Thus, LCR is assumed to have improved the recognition rate by using the discriminative ability from local manifolds of input images. In fact, VLR and HR images are not only different in resolution, but also in manifold structures. Therefore, selecting a best performance parameter for LCR is significance in experiment. Subsequently, we design an experiment to confirm the roles of LCR for testing the HR and LR manifold structures.

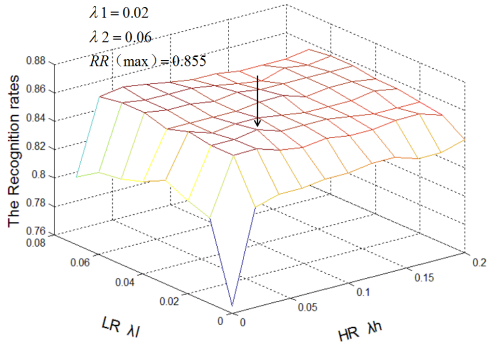

Fig. 7. The Recognition rate values versus different λs.

We employ the AR database experiments as an example. We show the VLR and HR locality constraint weights and their recognition performance in Fig. 7. λ_1 and λ_2 indicate VLR and HR control parameters, which are updated in 0.02 intervals. When $\lambda_1 = 0.02$ and $\lambda_2 = 0.06$ the recognition rate (RR) reaches the optimum value. General, discriminative information of HR image is naturally much more than its LR version. Thus, the locality-constrained term in HR should be more credible, as indicated by the greater contribution weights. In the same manner, we plot the peak signal-to-noise ratio (PSNR) values of different λ settings in Fig. 8. From this figure, we see that the parameters λ_1 and λ_2 have a obvious impact on PSNR performance. When $\lambda_1 = 0.01$ and $\lambda_2 = 0.06$ the PSNR (dB) score reached its best performance. For both recognition and SR tasks, the HR images provide stable manifold structure priors, thereby enabling the same locality balance parameters. However, slight differences can be observed in the LR image space. Nonetheless, given that the semi-coupled dictionary learning scheme drops the locality terms during training, the slight difference in LR balance parameter is acceptable. The process of estimating the balance parameters is described in the appendix.

D. Recognition rate

We fully compare our approach with a number of state-of-the-art recognition methods. Here, two settings are used in the experiments: first setting is using vision-based recognition algorithm. Super-resolved HR images are used as testing data which is cascaded with several popular classifier engines, i.e., SRC [58], CRC [59] and PCANET [60] (The recognition performance is shown in Table I). The second one is a typical resolution-robust VLR face recognition method that only uses LR images as testing inputs, including FRH method [37] and CDMMA [25]. Among the list, PCANET [60] has been

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

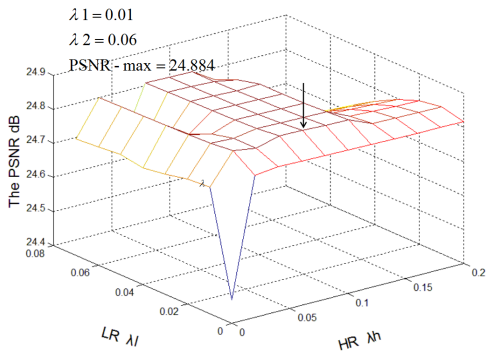

Fig. 8. The PSNR values versus different λs.

TABLE I
RECOGNITION PERFORMANCE OF SOME VISION-BASED FACE RECOGNITION METHODS USING AR DATABASE AND CMU PIE DATABASES.

Methods	SRC		CRC		PCANET	
	AR	CMU	AR	CMU	AR	CMU
HR	0.9450	0.9421	0.9550	0.9497	0.9575	0.9309
VLR	0.8200	0.8872	0.8280	0.8827	0.0275	0.0564
BIC	0.3000	0.5672	0.3550	0.4576	0.4400	0.4972
LSR	0.7875	0.8598	0.8180	0.8832	0.8075	0.8121
SRSR	0.7850	0.8765	0.7750	0.8791	0.7575	0.8385
LCR	0.7785	0.8644	0.7970	0.8756	0.7650	0.8121

known as a baseline of deep-learning based face recognition algorithm. We cascaded VDSR [56] with the CRC classifier and the SRC classifier using the same configuration to test the recognition performance in Table II. We fine-tune the selected classifiers at their best performance. We average five-times performance score in Tables I and II. When the input VLR images are super-resolved into the HR images using traditional classifiers, findings showed that CRC is better than other methods in term of recognition rates. Compared with VLR face image recognition, performance of simply feeding the super-resolved images into traditional classifiers is even lower than the original LR version. Moreover, vision-based hallucination does not seem to contribute to recognition. The results has similar performance reports with the previous work [44]. Feature-based VLR recognition, for an example: FRH, exhibits higher recognition rates than SR-based algorithms. In the other hand, as shown in Tables I and II, SLR outperforms than the FRH [37] and CDMMA [25], as shown by improvements of 6.00% and 4.15% using the AR database and 2.64% and 1.32% using the CMU database, respectively. Here CDMMA uses coupled-learning scheme too. Above results confirm that SLR has stronger discriminative ability than its competitors. The proposed semi-coupled learning scheme improves the recognition rate score by fully utilizing HR and VLR manifold constraints.

E. Face hallucination

In this sub-section, we assess the subjective and objective image qualities of all competitors, including bicubic interpolation (BIC), LSR [9], SRSR [49], LCR [12], CNE [41], FRH [37] and VDSR [56].

TABLE II
RECOGNITION RATE OF SOME STATE-OF-THE-ART RESOLUTION-ROBUST-BASED FACE RECOGNITION USING AR AND CMU PIE DATABASES.

Databases	FRH [37]	CDMMA [25]	VDSR+CRC	VDSR+SRC	SLR
AR	0.7950	0.8135	0.8125	0.8175	0.8550
CMU	0.8893	0.9025	0.9076	0.9072	0.9157

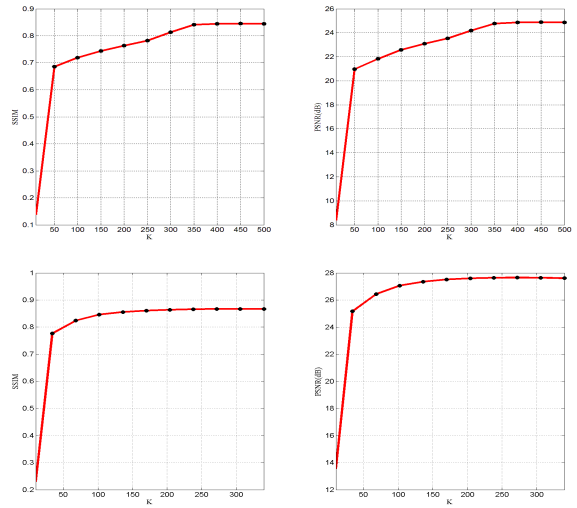

Fig. 9. The SSIM and PSNR values versus different K values. First row indicates results using AR database. Second row indicates results using CMU PIE database.

1) Quantity of Neighbors : We test the amount of K neighbors on the basis of two objectives, namely, to save on computing cost and to boost subjective image quality. For the best performance, we fine-tuned K with different settings. The SSIM and PSNR of the AR and CMU PIE databases are shown in Fig. 9. For the AR database, when $K = 350$, the SSIM and PSNR achieved the optimum. For the CMU PIE database, when $K = 200$, the SSIM and PSNR obtained their best performance. Findings indicate that the shrinking scheme of the proposed approach boosts its performance by using the locality prior from the training samples.

2) Objective reconstruction quality: Here three objective quality measures, namely, PSNR, structural similarity (SSIM) and feature similarity (FSIM), are used to evaluate the performances of the different algorithms. The results are shown in Tables III and IV.

We average all of the testing results and list them in Table III and IV to highlight the promoted performance compared with those of competitors. Here, VDSR gets the best performances on PSNR and SSIM on the AR database and second-best performances on the CMU PIE databases respectively. However, when VDSR results are fed into the CRC recognition engine (Table II), its recognition performance is lower than the proposed method in terms of both on AR and CMU PIE databases. In the same time, FSIM scores of SLR are slightly higher than VDSR. This phenomenon demonstrates that features have a dominant role in recognition than visual results. SLR yields the best performances comparing with other representation-based approaches without VDSR.

3) Time complexity: We implement the SLR algorithm in Matlab with hardware configuration: Intel Core i5-6300HQ

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

TABLE III
COMPARATIVE AVERAGES OF PSNR ,SSIM AND FSIM USING THE AR DATABASE.

Methods	BIC	SRSR	LSR	LCR	CNE	FRH	VDSR	Ours
PSNR(dB)	20.467	24.273	24.076	24.439	24.655	24.729	25.283	24.884
SSIM	0.563	0.829	0.818	0.829	0.835	0.844	0.857	0.845
FSIM	0.785	0.908	0.903	0.909	0.927	0.912	0.913	0.914

TABLE IV
COMPARATIVE AVERAGES OF PSNR ,SSIM AND FSIM USING CMU PIE DATABASE.

Methods	BIC	SRSR	LSR	LCR	CNE	FRH	VDSR	Ours
PSNR(dB)	22.031	26.511	25.709	26.985	27.288	27.186	27.405	27.645
SSIM	0.561	0.840	0.817	0.858	0.848	0.854	0.857	0.867
FSIM	0.780	0.908	0.904	0.913	0.905	0.914	0.915	0.921

CPU @2.30GHz, 8 GBytes RAM for experiments. As shown in Fig. 10, VDSR has best performance on running time testing (0.006sec/image) with GPU acceleration, but its training time is much longer than other methods, and its optimization process relies on GPU devices, so it is unfair to directly compare time effectiveness with SLR. SLR takes slightly more running-time than BIC and LSR, but it has much higher PSNR scores both on AR and CMU databases than BIC and LSR. On the other hand, SLR has better running time and PSNR performances than FRH, LCR, SRSR and CNE. In short, the running-time of different algorithms confirms the time effectiveness of SLR.

4) Visual results: The visual results of using the different algorithms are shown in Figs. 11 and 12. The HR images have 32×24 pixels, and thus, they are somewhat unclear. By contrast, the VLR images with 8×6 pixels result in blurriness that prevents the effective gathering of facial information. Except for BIC, all the other SR algorithms achieve smooth results. Specifically, the results of the proposed approach contains more details (e.g., edges of nose, eye, and mouth in magnified version) compared with those of the other methods. Although the VDSR obtained higher PSNR and SSIM performance, the visual results seem to be more blurred/smoother details compared with those from our method. This finding indicates that the visual reconstruction performance of SLR is competitive with other competitive algorithms.

V. CONCLUSION

We propose a novel semi-coupled locality-constrained representation algorithm to improve the discriminative and reconstructive abilities for VLR image in this work. Semi-coupled learning scheme fully uses manifold consistency that revises the feature representation capabilities. Then we conduct comprehensive experiments on VLR face recognition and hallucination using AR and CMU PIE databases, and results confirm the effectiveness of the proposed method over several state-of-the-art SR and recognition methods. In the future, the recognition performance should be further boosted by designing novel deep feature representation schemes.

VI. APPENDIX

Although locality constraint leads to smooth sparsity which boosts its performance, the regularization parameter λ is still hard to estimate. Vanilla solution to estimate λ is to fine tune the best performance parameter. In this paper, we use a maximum a posteriori probability (MAP) method to estimate this parameter. Take how to estimate LR locality constraint parameter λ_1 for example, for an input $\boldsymbol{x}_l^i$, D_l is LR dictionary, then LCR objective function is defined as following

$$\arg\min_{\boldsymbol{\alpha}_l^i}\{\left\|\boldsymbol{x}_l^i - D_l\boldsymbol{\alpha}_l^i\right\|_2^2 + \lambda_1 \left\|\boldsymbol{l}_l^i \circ \boldsymbol{\alpha}_l^i\right\|_2^2\}, s.t. \mathbf{1}^T\boldsymbol{\alpha} = 1. \tag{19}$$

From statistical point of view, input images $\boldsymbol{x}_l^i$ is observed variable, the task is to estimate unobservable variable $\boldsymbol{\alpha}_l^i$. Then objective function is defined as:

$$p(\boldsymbol{\alpha}_l^i|\boldsymbol{x}_l^i) = \frac{p(\boldsymbol{\alpha}_l^i)p(\boldsymbol{x}_l^i|\boldsymbol{\alpha}_l^i)}{p(\boldsymbol{x}_l^i)}, \tag{20}$$

where $p(\boldsymbol{x}_l^i|\boldsymbol{\alpha}_l^i)$ represents a posterior probability, in general, the image noise is assume as Gaussian distribution with zero mean and σ variance, then:

$$p(\boldsymbol{x}_l^i|\boldsymbol{\alpha}_l^i) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}||\boldsymbol{x}_l^i - D_l\boldsymbol{\alpha}_l^i||_2^2), \tag{21}$$

usually, representation coefficients are assumed as Gaussian distribution with zero mean and σ_r variance too. Then

$$p(\boldsymbol{\alpha}_l^i) = \frac{1}{\sqrt{2\pi}\sigma_r}\exp(-\frac{1}{2\sigma_r}||l_l^i \circ \boldsymbol{\alpha}||_2^2), \tag{22}$$

here l_l^i is a constant for locality metric. For a given observation variable $\boldsymbol{x}_l^i$, the formula for representation coefficient using MAP is given by:

$$\begin{aligned}\boldsymbol{\alpha}_l^i &= \arg\max_{\boldsymbol{\alpha}_l^i} \log p(\boldsymbol{\alpha}_l^i|\boldsymbol{x}_l^i) \\ &= \arg\max_{\boldsymbol{\alpha}_l^i}(logp(\boldsymbol{x}_l^i|\boldsymbol{\alpha}_l^i) + logp(\boldsymbol{\alpha}_l^i)).\end{aligned} \tag{23}$$

Formula (19) and (20) are combined into formula (21), then we have,

$$\begin{aligned}\boldsymbol{\alpha}_l^i = \arg\max_{\boldsymbol{\alpha}_l^i}(&log\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}||\boldsymbol{x}_l^i - D_l\boldsymbol{\alpha}_l^i||_2^2 \\ &+log\frac{1}{\sqrt{2\pi}\sigma_r} - \frac{1}{2\sigma_r^2}||l_l^i \circ \boldsymbol{\alpha}_l^i||_2^2.\end{aligned} \tag{24}$$

Since the first and third terms in formula (22) are constant, they always do not affect the probability of $\boldsymbol{\alpha}_l^i$, then we have

$$\boldsymbol{\alpha}_l^i = \arg\min_{\boldsymbol{\alpha}_l^i}(\frac{1}{2\sigma^2}||\boldsymbol{x}_l^i - D_l\boldsymbol{\alpha}_l^i||_2^2 + \frac{1}{2\sigma_r^2}||l_l^i \circ \boldsymbol{\alpha}_l^i||_2^2). \tag{25}$$

We compare above formula (24) with formula (18), it is easy to get:

$$\lambda_1 = \frac{\sigma^2}{\sigma_r^2} \tag{26}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2872761, IEEE Access

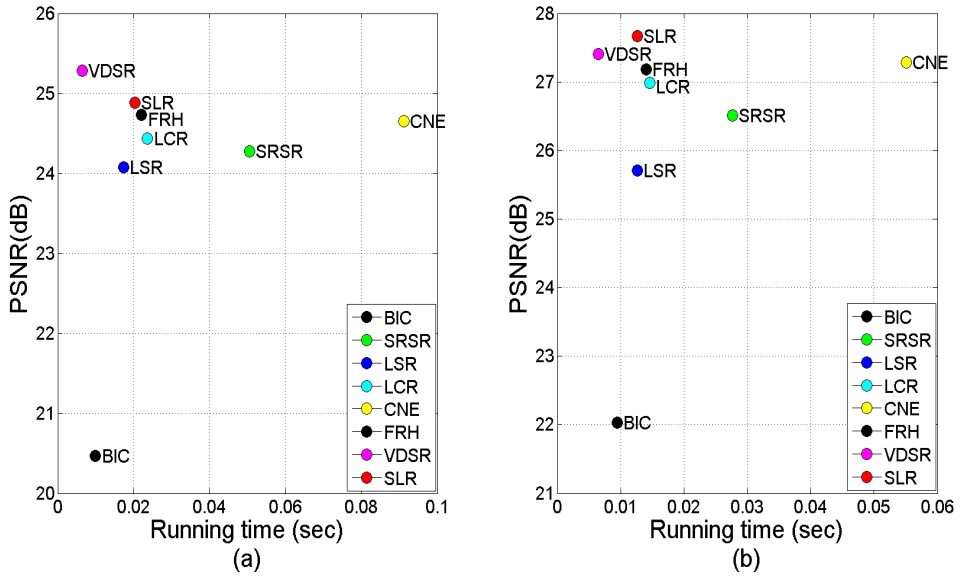

Fig. 10. Running time (seconds per testing image) of different SR algorithms using AR database(a) and CMU database(b), the scaling factor is 4, X axis indicates running time of algorithms and Y axis represents the PSNR scores.

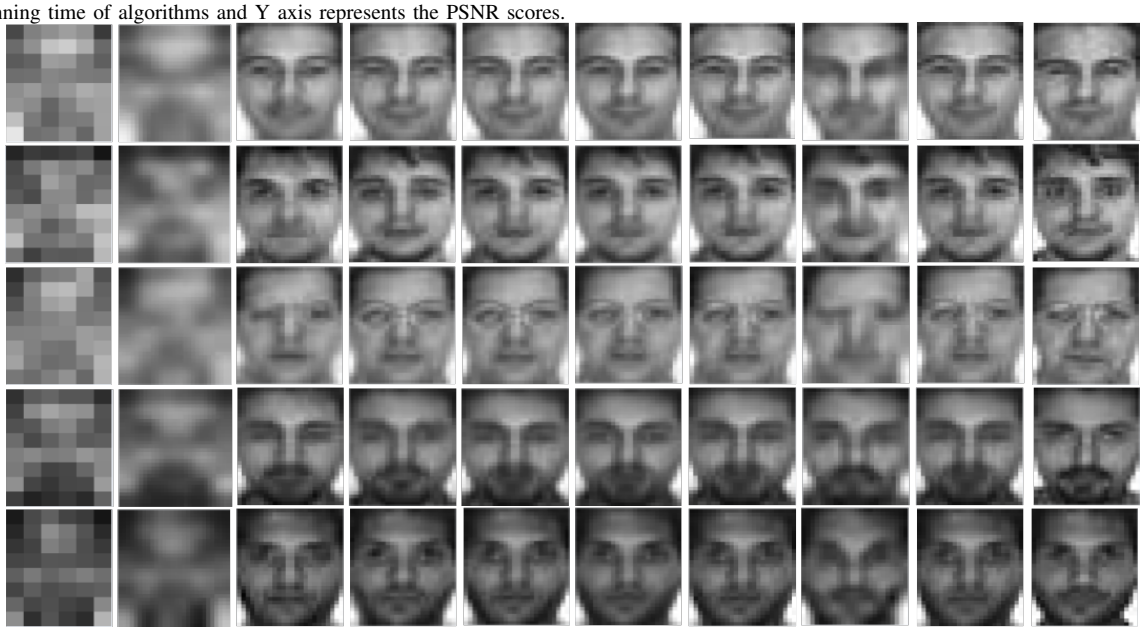

Fig. 11. Experimental results on image SR using AR database (scaling factor: 4). From left to right: LR image, Bicubic, LSR, SRSR, LCR, CNE, FRH, VDSR, the proposed SLR method and the HR ground-truth.

Where σ is noise variance of the observed image and σ_r is variance of representation coefficients. Here, we use same method (MAP) to estimate LR and HR locality term parameters λ_1 and λ_2 to perfectly exploit their locality constraints.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (61502354, 61671332), the Natural Science Foundation of Hubei Province of China (2018ZYY-D059, 2012FFA099, 2012FFA134, 2013CF125, 2014CFA130, 2015CFB451), Scientific Research Foundation of Wuhan Institute of Technology (K201713), Graduate Innovative Fund of Wuhan Institute of Technology (CX2017071), and in part by "The Pearl River Talent Recruitment Program Innovative and Entrepreneurial Teams in 2017" under grant No. 2017ZT07X152.

REFERENCES

[1] W. Zou and P. Yuen, "Very low resolution face recognition problem," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 327–340, Jan 2012.

[2] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 9, pp. 1167–1183, Sep 2002.

[3] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Processing*, vol. 128, pp. 389 – 408, 2016.

[4] Z. Shao, J. Cai, and Z. Wang, "Smart monitoring cameras driven intelligent processing to big surveillance video data," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2018.

[5] Z. Wang, Z. Miao, Q. M. J. Wu, Y. Wan, and Z. Tang, "Low-resolution face recognition: A review," *Visual Computer*, vol. 30, no. 4, pp. 359–386, 2014.

[6] T. K. Simon Baker, "Hallucinating Faces," in *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) 2000*, 2000, pp. 83–88.

Fig. 12. Experimental results on image SR using CMU database (scaling factor: 4). From left to right: LR image, Bicubic, LSR, SRSR, LCR, CNE, FRH, VDSR, the proposed SLR method and the HR ground-truth.

[7] X. W. X. Wang and X. T. X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, 2005.

[8] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 1, no. 3, pp. I–I, 2004.

[9] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch." *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.

[10] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *Signal Processing Letters, IEEE*, vol. 18, no. 6, pp. 367–370, June 2011.

[11] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in *2012 IEEE International Conference on Multimedia and Expo*, 2012, pp. 212–217.

[12] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1268–1281, 2014.

[13] J. Shi, X. Liu, and C. Qi, "Global consistency, local sparsity and pixel correlation: A unified framework for face hallucination," *Pattern Recognition*, vol. 47, no. 11, pp. 3520–3534, 2014.

[14] J. Shi and C. Qi, "Kernel-based face hallucination via dual regularization priors," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1189–1193, 2015.

[15] G. Gao, X. Y. Jing, P. Huang, Q. Zhou, S. Wu, and D. Yue, "Locality-constrained double low-rank representation for effective face hallucination," *IEEE Access*, vol. 4, no. 99, pp. 8775–8786, 2016.

[16] T. Lu, Z. Xiong, Y. Zhang, B. Wang, and T. Lu, "Robust face super-resolution via locality-constrained low-rank representation," *IEEE Access*, vol. 5, no. 99, pp. 13 103 – 13 117, 2017.

[17] T. Lu, L. Pan, H. Wang, Y. Zhang, B. Wang, and Z. Xiong, "Face hallucination using deep collaborative representation for local and non-local patches," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4.

[18] T. Lu, L. Pan, J. Jiang, Y. Zhang, and Z. Xiong, "Dlml: Deep linear mappings learning for face super-resolution with nonlocal-patch," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 1362–1367.

[19] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[20] T. Lu, H. Wang, Z. Xiong, J. Jiang, Y. Zhang, H. Zhou, and Z. Wang, "Face hallucination using region-based deep convolutional networks," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 1657–1661.

[21] J. Shi, X. Liu, Y. Zong, C. Qi, and G. Zhao, "Hallucinating face image by regularization models in high-resolution feature space," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2980–2995, 2018.

[22] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3991–4002, 2017.

[23] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "Srlsp: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 27–40, 2017.

[24] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, Jan 2010.

[25] J. Jiang, R. Hu, Z. Wang, and Z. Cai, "Cdmma: Coupled discriminant multi-manifold analysis for matching low-resolution face images," *Signal Processing*, vol. 124, pp. 162–172, 2016.

[26] J. Wang, S. Zhu, and Y. Gong, "Resolution-invariant image representation for content-based zooming," in *2009 IEEE International Conference on Multimedia and Expo*, June 2009, pp. 918–921.

[27] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 4792–4800.

[28] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain, *Nighttime Face Recognition at Long Distance: Cross-Distance and Cross-Spectral Matching*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 708–721.

[29] J. Shi and C. Qi, "From local geometry to global structure: Learning latent subspace for low-resolution face image recognition," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 554–558, 2014.

[30] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1706–1721, 2014.

[31] J. Ma, J. Zhao, H. Guo, J. Jiang, H. Zhou, and Y. Gao, "Locality preserving matching," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2017, pp. 4492–4498.

[32] Z. Yu, H. Zhou, and C. Li, "Fast non-rigid image feature matching for agricultural uav via probabilistic inference with regularization techniques," *Computers and Electronics in Agriculture*, vol. 143, pp. 79–89, 2017.

[33] T. Lu, Y. Guan, Y. Zhang, S. Qu, and Z. Xiong, "Robust and efficient face recognition via low-rank supported extreme learning machine," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 11 219–11 240, 2018.

[34] P. Hennings-Yeomans, S. Baker, and B. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[35] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer, "Pose-robust recognition of low-resolution face images," *Pattern Analysis & Machine Intelligence IEEE Transactions on*, vol. 35, pp. 3037–3049, 2013.

[36] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, Oct 2012.

[37] M. C. Yang, C. P. Wei, Y. R. Yeh, and Y. C. F. Wang, "Recognition at a long distance: Very low resolution face recognition and hallucination," in *Biometrics (ICB), 2015 International Conference on*, 2015.

[38] H. Huang and H. He, "Super-resolution method for face recognition using nonlinear mappings on coherent features," *Neural Networks, IEEE Transactions on*, vol. 22, no. 1, pp. 121–130, Jan 2011.

[39] M. Jian and K. M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1761–1772, Nov 2015.

[40] K. Su, Q. Tian, Q. Xue, N. Sebe, and J. Ma, "Neighborhood issue in single-frame image super-resolution," in *2005 IEEE International Conference on Multimedia and Expo*, July 2005, pp. 4 pp.–.

[41] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1674–1684, Sept 2016.

[42] R. Abiantun, M. Savvides, and B. V. K. V. Kumar, "How low can you go? low resolution face recognition study using kernel correlation feature analysis on the frgcv2 dataset," in *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, Sept 2006, pp. 1–6.

[43] D. Dai, Y. Wang, Y. Chen, and L. V. Gool, "Is image super-resolution helpful for other vision tasks?" in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–9.

[44] X. Xu, W. Liu, and L. Li, "Face hallucination: How much it can improve face recognition," in *Control Conference (AUCC), 2013 3rd Australian*, 2013, pp. 93–98.

[45] X. Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 695–704.

[46] S. Wang, L. Zhang, L. Y., and Q. Pan, "Semi-coupled dictionary learning with applications in image super-resolution and photo-sketch synthesis," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012.

[47] T. Lu, W. Yang, Y. Zhang, X. Li, and Z. Xiong, "Very low-resolution face recognition via semi-coupled locality-constrained representation," in *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, Dec 2016, pp. 362–367.

[48] J. Yu, Y. Rui, and D. Tao, "Click prediction for web image reranking using multimodal sparse coding." *Image Processing IEEE Transactions on*, vol. 23, no. 5, pp. 2019–2032, 2014.

[49] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[50] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face Hallucination Via Weighted Adaptive Sparse Regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 802–813, 2014.

[51] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 3360–3367.

[52] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends. Comput. Graph. Vis.*, vol. 8, no. 2-3, pp. 85–283, Dec. 2014.

[53] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognition*, vol. 47, no. 9, pp. 2794–2806, 2014.

[54] A. M. Martinez and A. C. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, Feb 2001.

[55] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, September 2008.

[56] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, June 2016.

[57] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[58] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[59] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *ICCV*, Nov 2011, pp. 471–478.

[60] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, Dec 2015.