Identity Aware Synthesis for Cross Resolution Face Recognition

Maneet Singh, Shruti Nagpal, Mayank Vatsa, Richa Singh, and Angshul Majumdar
IIIT-Delhi, India
{maneets, shrutin, mayank, rsingh, angshul}@iiitd.ac.in

Abstract

Enhancing low resolution images via super-resolution or synthesis algorithms for cross-resolution face recognition has been well studied. Several image processing and machine learning paradigms have been explored for addressing the same. In this research, we propose Synthesis via Hierarchical Sparse Representation (SHSR) algorithm for synthesizing a high resolution face image from a low resolution input image. The proposed algorithm learns multi-level sparse representation for both high and low resolution gallery images, along with identity aware dictionaries and a transformation function between the two representations for face identification scenarios. With low resolution test data as input, a high resolution test image is synthesized using the identity aware dictionaries and transformation, which is then used for face recognition. The performance of the proposed SHSR algorithm is evaluated on four datasets, including one real world dataset. Experimental results and comparison with seven existing algorithms demonstrate the efficacy of the proposed algorithm in terms of both face identification and image quality measures.

1. Introduction

Group images are often captured from a distance, in order to capture multiple people in the image. In such cases, the resolution of each face image is relatively small, thereby resulting in errors during automated tagging or recognition. Similarly, in surveillance and monitoring applications, cameras are often designed to cover the maximum field of view, thus limiting the size of face images captured, especially for individuals at a distance. These low resolution images are often used to match against high resolution images, e.g., profile images on social media or mugshot images captured by law enforcement. In such scenarios, the resolution gap between the two may lead to incorrect results. This task of matching a low resolution input image against a database of high resolution images is referred to as cross resolution face recognition.

Several researchers have shown that the performance of state-of-the-art (SOTA) face recognition algorithms reduces while matching cross-resolution face images [4, 36, 37]. In order to overcome this limitation, an intuitive approach is to generate a high resolution image for the given low resolution input, which can be provided as input to the face recognition engine. Figure 1 shows sample real world images captured minutes before the Boston Bombing (2013). Since the person of interest is at distance, the face captured is thus of low resolution. Performing bicubic interpolation to obtain a high resolution image results in an image suffering from blur and of poor quality. With the aim of high recognition performance, the generated high resolution image should have good quality while preserving the identity of the subject. As elaborated in the next subsection, while there exist multiple synthesis or super resolution techniques, we hypothesize that utilizing a (domain) face-specific, recognition-oriented model for face synthesis will result in improved recognition performance, especially for close-set recognition scenarios. To this effect, this work presents a novel domain specific identity aware *Synthesis via Hierarchical Sparse Representation (SHSR)* algorithm for synthesizing a high resolution face image from a given low resolution input image.

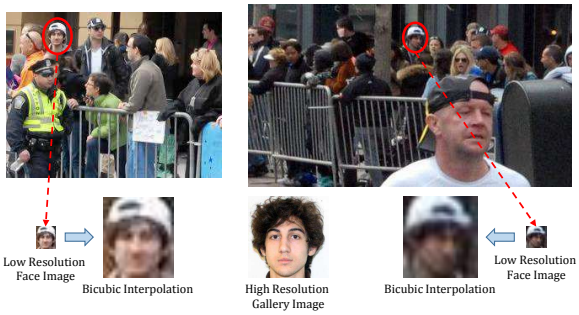

Figure 1: Images (taken from the Internet) captured minutes before Boston Marathon Bombing, 2013, of suspect Dzhokhar Tsarnaev (circled). The resolution of the circled image is less than 24×24, which is interpolated to (96×96).

1.1. Literature Review

In literature, different techniques have been proposed to address the problem of cross resolution face recogni-

tion. These can broadly be divided into transformation based techniques and non-transformation based techniques. Transformation based techniques address the resolution difference between images by explicitly introducing a transformation function either at the image or at the feature level. Non-transformation techniques propose to extract/learn resolution invariant features or classifiers, in order to address the resolution variations [4, 12, 36]. In 2013, Wang *et al.* [37] present an exhaustive review of the proposed techniques for addressing cross resolution face recognition.

Peleg and Elad [27] propose a statistical model that uses Minimum Mean Square Error estimator on high and low resolution image pair patches for prediction. Lam [19] propose a Singular Value Decomposition based approach for super resolving low resolution face images. Researchers have also explored the domain of representation learning to address the problem of cross resolution face recognition. Yang *et al.* [38] propose learning dictionaries for low and high resolution image patches jointly, followed by learning a mapping between the two. Yang *et al.* [39] propose a Sparse Representation-based Classification approach in which the face recognition and hallucination constraints are solved simultaneously. Gu *et al.* [18] propose convolutional sparse coding where an image is divided into patches and filters are learned to decompose a low resolution image into features. A mapping is learned to predict high resolution feature maps from the low resolution features. Mundunuri and Biswas [25] propose a multi-dimensional scaling and stereo cost technique to learn a common transformation matrix for addressing the resolution variations.

A parallel area of research is that of super-resolution, where research has focused on obtaining a high resolution image from a given low resolution image, with the objective of maintaining/improving the visual quality of the input [3, 28, 34, 35]. There has been significant advancement in the field of super-resolution over the past several years including recent representation learning architectures [5, 8, 9, 22, 32] being proposed for the same. It is important to note that while such techniques can be utilized for addressing cross resolution face recognition, however, they are often not explicitly trained for face images, or for providing recognition-oriented results.

1.2. Research Contributions

This research focuses on cross resolution face recognition by proposing a recognition-oriented image synthesis algorithm, capable of handling large magnification factors. We propose a hierarchical sparse representation based transfer learning approach, termed as Synthesis via Hierarchical Sparse Representation (SHSR). The proposed identity aware synthesis algorithm can be incorporated as a preprocessing module prior to any existing face recognition engine to enhance the resolution of a given low resolution input. In order to ensure recognition-oriented synthesis, the proposed model is trained using a gallery database containing a single image per subject. The results are demonstrated with four datasets in terms of face identification accuracies with existing face recognition models and no-reference image quality measure of the synthesized images.

2. Synthesis via Hierarchical Sparse Representation (SHSR)

Dictionary learning algorithms have an inherent property of representing a given sample as a sparse combination of its basis functions [29]. This property is utilized in the proposed SHSR algorithm to synthesize a high resolution image from a given low resolution input. The proposed model learns a transformation between the representations of low and high resolution images. Further, motivated by the abstraction capabilities of deep learning, we propose to learn the transformation between deeper levels of representation. Unlike traditional dictionary learning algorithms, we propose to learn the transformation at higher levels of representation. This leads to the key contribution of this work: *Synthesis via Hierarchical Sparse Representation*, a transfer learning approach for synthesizing a high resolution image for a given low resolution input.

2.1. Preliminaries

Let $\mathbf{X} = [\mathbf{x}^1|\mathbf{x}^2|...|\mathbf{x}^n]$ be the training data with n samples. Dictionary learning algorithms learn a *dictionary* ($\mathbf{D}$) and *sparse representations* ($\mathbf{A}$) using data ($\mathbf{X}$). The objective function of dictionary learning is written as:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{n}\sum_{i=1}^{n}\left(\left\|\mathbf{x}^i - \mathbf{D}\boldsymbol{\alpha}^i\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}^i\right\|_1\right) \quad (1)$$

where, $\mathbf{A} = [\boldsymbol{\alpha}^1|\boldsymbol{\alpha}^2|....|\boldsymbol{\alpha}^n]$ are the sparse codes, $\|\cdot\|_1$ represents ℓ_1-norm, and λ is the regularizing constant that governs the weight given to induce sparsity in the representations. In Eq. 1, the first term minimizes the reconstruction error of the training samples, and the second term is a regularization term on the sparse codes.

In literature, researchers have extended a single level dictionary to a multi-level or hierarchical dictionary to learn multiple levels of representation of the given data [26, 30, 31]. A $k-$level hierarchical dictionary learns k dictionaries $\mathbf{D} = \{\mathbf{D}^1, ..., \mathbf{D}^k\}$ and sparse coefficients $\mathbf{A} = \{\mathbf{A}^1, ..., \mathbf{A}^k\}$ for a given input $\mathbf{X}$:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{n}\sum_{i=1}^{n}\left(\left\|\mathbf{x}^i - \mathbf{D}^1...\mathbf{D}^k\boldsymbol{\alpha}^{k,i}\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}^{k,i}\right\|_1\right) \quad (2)$$

where, $\alpha^{k,i}$ corresponds to the k^{th} level representation of the i^{th} sample. This architecture is analogous to deep learning techniques, where deeper layers of feature learning enhance the level of abstraction learned by the network, thereby

learning meaningful latent variables. A two layer hierarchical dictionary learning model is formulated as follows:

$$\min_{\substack{\mathbf{D}^1,\mathbf{D}^2,\\ \mathbf{A}^1,\mathbf{A}^2}} \frac{1}{n}\sum_{i=1}^{n}\Big(\left\|\mathbf{x}^i-\mathbf{D}^1\mathbf{D}^2\boldsymbol{\alpha}^{2,i}\right\|_2^2+\lambda_1\left\|\boldsymbol{\alpha}^{1,i}\right\|_1+\lambda_2\left\|\boldsymbol{\alpha}^{2,i}\right\|_1\Big) \tag{3}$$

here, $\boldsymbol{\alpha}^{1,i} = \mathbf{D}^2\boldsymbol{\alpha}^{2,i}$ such that $\mathbf{A}^1 = [\boldsymbol{\alpha}^{1,1}|\boldsymbol{\alpha}^{1,2}|...|\boldsymbol{\alpha}^{1,n}]$. the above equation can be modeled as a step-wise optimization of the following two equations:

$$\min_{\mathbf{D}^1,\mathbf{A}^1}\frac{1}{n}\sum_{i=1}^{n}\Big(\left\|\mathbf{x}^i-\mathbf{D}^1\boldsymbol{\alpha}^{1,i}\right\|_2^2+\lambda_1\left\|\boldsymbol{\alpha}^{1,i}\right\|_1\Big) \tag{4}$$

$$\min_{\mathbf{D}^2,\mathbf{A}^2}\frac{1}{n}\sum_{i=1}^{n}\Big(\left\|\boldsymbol{\alpha}^{1,i}-\mathbf{D}^2\boldsymbol{\alpha}^{2,i}\right\|_2^2+\lambda_2\left\|\boldsymbol{\alpha}^{2,i}\right\|_1\Big) \tag{5}$$

The above two equations correspond to standard dictionary learning formulations, which can be optimized using existing techniques such as alternating minimization over the dictionary and representation [29].

2.2. SHSR Algorithm

In real world scenarios of surveillance or image tagging, the task is to match a low resolution test image (probe) to the database of high resolution images known as *gallery images*. Without loss of generality, we assume that the target comprises of high resolution gallery images while the source domain consists of low resolution images. In the proposed model, for low resolution face images $\mathbf{X}_l$, and high resolution face images $\mathbf{X}_h$, k-level hierarchical dictionaries are learned in both source ($\mathbf{G_L} = \{\mathbf{G}_l^1, ..., \mathbf{G}_l^k\}$) and target domain ($\mathbf{G_H} = \{\mathbf{G}_h^1, ..., \mathbf{G}_h^k\}$). It is important to note that the dictionaries are generated using the pre-acquired gallery images. Corresponding sparse representations, $\mathbf{A_L} = \{\mathbf{A}_l^1, ..., \mathbf{A}_l^k\}$ and $\mathbf{A_H} = \{\mathbf{A}_h^1, ..., \mathbf{A}_h^k\}$ are also learned for all k levels, where $\mathbf{A}_h^k = [\boldsymbol{\alpha}_h^{k,1}|\boldsymbol{\alpha}_h^{k,2}|...|\boldsymbol{\alpha}_h^{k,n}]$ are the representations learned corresponding to the k^{th}-level high resolution dictionary and $\mathbf{A}_l^k = [\boldsymbol{\alpha}_l^{k,1}|\boldsymbol{\alpha}_l^{k,2}|...|\boldsymbol{\alpha}_l^{k,n}]$ are the representations learnt from the k^{th} level dictionary for the low resolution images. The proposed algorithm learns a transformation, $\mathbf{M}$, between $\mathbf{A}_h^k$ and $\mathbf{A}_l^k$. The optimization function for Synthesis via Hierarchical Sparse Representation, a k-level hierarchical dictionary is written as:

$$\begin{aligned}\min_{\substack{\mathbf{G_H},\mathbf{A_H},\\ \mathbf{G_L},\mathbf{A_L},\mathbf{M}}}\frac{1}{n}\sum_{i=1}^{n}\Big(&\left\|\mathbf{x}_h^i-\mathbf{G}_h^1...\mathbf{G}_h^k\boldsymbol{\alpha}_h^{k,i}\right\|_2^2+\sum_{j=1}^{k}(\lambda_j\left\|\boldsymbol{\alpha}_{\boldsymbol{h}}^{j,i}\right\|_1)\\&+\left\|\mathbf{x}_l^i-\mathbf{G}_l^1...\mathbf{G}_l^k\boldsymbol{\alpha}_l^{k,i}\right\|_2^2+\sum_{j=1}^{k}(\lambda_j\left\|\boldsymbol{\alpha}_{\boldsymbol{l}}^{j,i}\right\|_1)\\&+\lambda_M\left\|\boldsymbol{\alpha}_h^{k,i}-\mathbf{M}\boldsymbol{\alpha}_l^{k,i}\right\|_2^2\Big)\end{aligned} \tag{6}$$

where, λ_j are regularization parameters governing the amount of sparsity in the learned representations of the j^{th} layer, while λ_M is the regularization constant for learning the transformation function. $\mathbf{G_H}$ and $\mathbf{G_L}$ correspond to the hierarchical dictionaries learned for the high and low resolution gallery images, respectively. The first two terms are responsible for learning the dictionaries and representations of high resolution face images, while the next two terms correspond to feature learning for low resolution face images. The last term contains the transformation between the deepest features of both the resolutions. Therefore, the SHSR algorithm learns multiple levels of dictionaries and corresponding representations for low and high resolution face images, along with a transformation between the features learned at the last layer.

2.2.1 Training SHSR Algorithm

Without loss of generality, training of the proposed SHSR algorithm is explained with $k = 2$ (shown in Figure 2). For a two level hierarchical dictionary, Eq. 6 can be written as:

$$\begin{aligned}\min_{\substack{\mathbf{G_H},\mathbf{G_L},\\ \mathbf{A_H},\mathbf{A_L},\mathbf{M}}}\frac{1}{n}\sum_{i=1}^{n}\Big(&\left\|\mathbf{x}_h^i-\mathbf{G}_h^1\mathbf{G}_h^2\boldsymbol{\alpha}_h^{2,i}\right\|_2^2+\lambda_1\left\|\boldsymbol{\alpha}_h^{1,i}\right\|_1\\&+\lambda_2\left\|\boldsymbol{\alpha}_h^{2,i}\right\|_1+\left\|\mathbf{x}_l^i-\mathbf{G}_l^1\mathbf{G}_l^2\boldsymbol{\alpha}_l^{2,i}\right\|_2^2+\lambda_1\left\|\boldsymbol{\alpha}_l^{1,i}\right\|_1\\&+\lambda_2\left\|\boldsymbol{\alpha}_l^{2,i}\right\|_1+\lambda_M\left\|\boldsymbol{\alpha}_h^{2,i}-\mathbf{M}\boldsymbol{\alpha}_l^{2,i}\right\|_2^2\Big)\end{aligned} \tag{7}$$

Since the number of variables in Eq. 7 is large, directly solving the optimization problem may provide inaccurate estimates. Therefore, greedy layer by layer training is applied. It is important to note that since there is a l_1-norm regularizer on the coefficients of the first and the second layer, the dictionaries $\mathbf{G}_h^1$ and $\mathbf{G}_h^2$ cannot be collapsed into one dictionary. A hierarchical dictionary of two levels (Eq. 3) is learned in two steps (Eq. 4, 5). Upon extending the formulation to k levels, it would require exactly k steps for optimization. The proposed SHSR algorithm (Eq. 6) builds upon the above and utilizes $k + 1$ steps based greedy layer-wise learning for a k-level dictionary. k steps are for learning representations using the dictionary architecture and the $k + 1^{th}$ step is for learning the transformation between the final representations. Therefore, Eq. 7 is solved using an independent three step approach: **(i)** learn first level source (low resolution) and target (high resolution) domain dictionaries, **(ii)** learn second level low and high resolution dictionaries, and **(iii)** learn a transformation between the final representations.

Using the concept in Eq. 3 - 5, in the **first step**, two separate level-1 (i.e. $k = 1$) dictionaries are learned from the given input data for the low resolution ($\mathbf{G}_l^1$) and high resolution ($\mathbf{G}_h^1$) face images independently. Given the training data consisting of low ($\mathbf{X}_l$) and high ($\mathbf{X}_h$) resolution face images, the following minimization is applied for the two

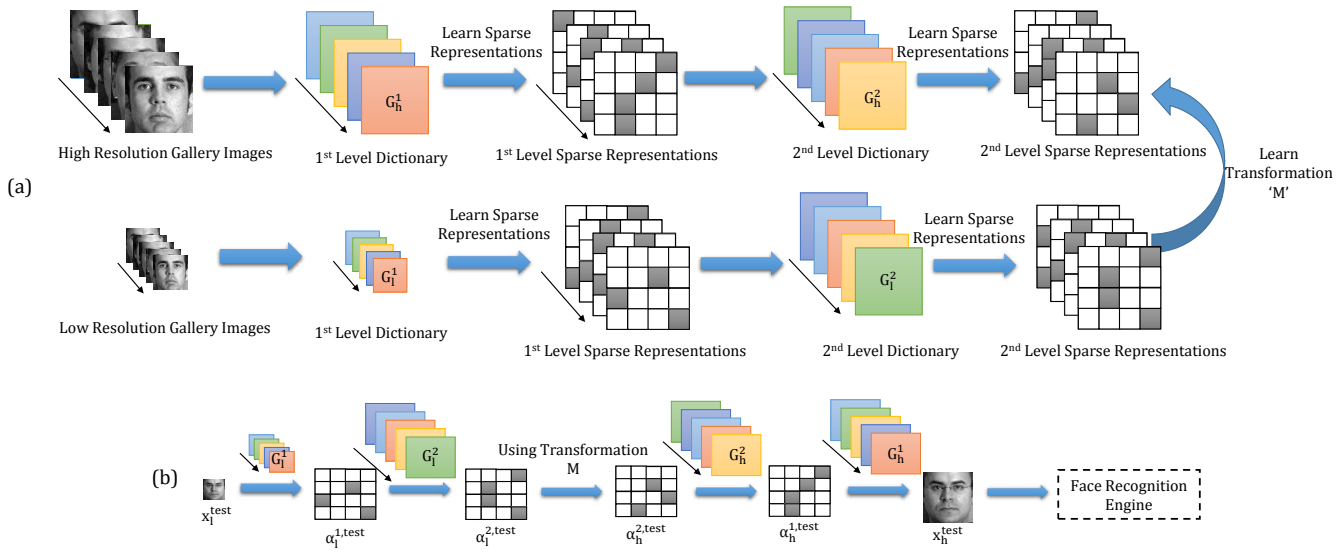

Figure 2: Synthesis via Hierarchical Sparse Representation algorithm for 2-level hierarchical dictionary. (a) shows the training of the model, while (b) illustrates the high resolution synthesis of a low resolution input.

domains respectively:

$$\min_{\mathbf{G}_l^l,\,\mathbf{A}_l^l} \frac{1}{n}\sum_{i=1}^{n} \left\|\mathbf{x}_l^i - \mathbf{G}_l^l\boldsymbol{\alpha}_l^{l,i}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\alpha}_l^{l,i}\right\|_1 \tag{8}$$

$$\min_{\mathbf{G}_h^l,\,\mathbf{A}_h^l} \frac{1}{n}\sum_{i=1}^{n} \left\|\mathbf{x}_h^i - \mathbf{G}_h^l\boldsymbol{\alpha}_h^{l,i}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\alpha}_h^{l,i}\right\|_1 \tag{9}$$

here, $\mathbf{A}_l^l = [\boldsymbol{\alpha}_l^{l,1}|\boldsymbol{\alpha}_l^{l,2}|...|\boldsymbol{\alpha}_l^{l,n}]$ and $\mathbf{A}_h^l = [\boldsymbol{\alpha}_h^{l,1}|\boldsymbol{\alpha}_h^{l,2}|...|\boldsymbol{\alpha}_h^{l,n}]$ refer to the level-1 sparse codes learned for the low and high resolution images, respectively. Each of the above two equations can be optimized independently using an alternating minimization dictionary learning technique over the dictionary and representation [23]. After this step, $\mathbf{G}_h^l, \mathbf{G}_l^l$ (dictionaries) and $\mathbf{A}_h^l, \mathbf{A}_l^l$ (representations) are obtained for the high and low resolution data.

In the **second step**, a hierarchical dictionary is created by learning the second level dictionaries ($\mathbf{G}_l^2, \mathbf{G}_h^2$) using the representations obtained from the first level ($\mathbf{A}_h^l$ and $\mathbf{A}_l^l$). That is, two separate dictionaries, one for low resolution images and one for high resolution images are learned using the representations obtained at the first level as input features. The equations for this step are written as follows:

$$\min_{\mathbf{G}_l^2,\,\mathbf{A}_l^2} \frac{1}{n}\sum_{i=1}^{n} \left\|\boldsymbol{\alpha}_l^{l,i} - \mathbf{G}_l^2\boldsymbol{\alpha}_l^{2,i}\right\|_2^2 + \lambda_2 \left\|\boldsymbol{\alpha}_l^{2,i}\right\|_1 \tag{10}$$

$$\min_{\mathbf{G}_h^2,\,\mathbf{A}_h^2} \frac{1}{n}\sum_{i=1}^{n} \left\|\boldsymbol{\alpha}_h^{l,i} - \mathbf{G}_h^2\boldsymbol{\alpha}_h^{2,i}\right\|_2^2 + \lambda_2 \left\|\boldsymbol{\alpha}_h^{2,i}\right\|_1 \tag{11}$$

here, $\mathbf{A}_l^2 = [\boldsymbol{\alpha}_l^{2,1}|\boldsymbol{\alpha}_l^{2,2}|...|\boldsymbol{\alpha}_l^{2,n}]$ is the final representation obtained for the low resolution images and $\mathbf{A}_h^2 = [\boldsymbol{\alpha}_h^{2,1}|\boldsymbol{\alpha}_h^{2,2}|...|\boldsymbol{\alpha}_h^{2,n}]$ refers to the representation obtained for the high resolution images. Similar to the previous step, the equations can be solved independently using alternating minimization over the dictionary and representations. After this step, $\mathbf{G}_l^2, \mathbf{G}_h^2, \mathbf{A}_l^2$ and $\mathbf{A}_h^2$ are obtained.

In order to synthesize from one resolution to another, the **third step** of the algorithm involves learning a transformation between the final representations of the two resolutions (i.e. $\mathbf{A}_l^2$ and $\mathbf{A}_h^2$). The following minimization is solved to obtain a transformation $\mathbf{M}$:

$$\min_{\mathbf{M}} \left\|\mathbf{A}_h^2 - \mathbf{M}\mathbf{A}_l^2\right\|_F^2 \tag{12}$$

The above equation is a least square problem with a closed form solution. After training, the dictionaries ($\mathbf{G}_l^1, \mathbf{G}_h^1, \mathbf{G}_l^2, \mathbf{G}_h^2$) and the transformation function ($\mathbf{M}$) are obtained which are used at test time.

2.2.2 Testing: Synthesizing a High Resolution Face Image from a Low Resolution Image

During testing, a low resolution test image, $\mathbf{x}_l^{test}$, is input to the algorithm. Using the trained gallery based dictionaries, $\mathbf{G}_l^l$ and $\mathbf{G}_l^2$, first and second level representations ($\boldsymbol{\alpha}_l^{l,test}, \boldsymbol{\alpha}_l^{2,test}$) are obtained for the given image:

$$\mathbf{x}_l^{test} = \mathbf{G}_l^l\boldsymbol{\alpha}_l^{l,test};\ \boldsymbol{\alpha}_l^{l,test} = \mathbf{G}_l^2\boldsymbol{\alpha}_l^{2,test} \tag{13}$$

The transformation function, $\mathbf{M}$, learned in Eq. 12, is then used to obtain the second level high resolution representation ($\boldsymbol{\alpha}_h^{2,test}$):

$$\boldsymbol{\alpha}_h^{2,test} = \mathbf{M}\boldsymbol{\alpha}_l^{2,test} \tag{14}$$

Table 1: Summarizing the characteristics of the training and testing partitions of the datasets used in experiments.

Dataset	Training Subjects	Training Images	Testing Subjects	Testing Images	Gallery Resolution	Probe Resolutions
CMU Multi-PIE [17]	100	200	237	474	96×96	8×8, 16×16, 24×24, 32×32, 48×48
CAS-PEAL [13]	500	659	540	705		
Real World Scenarios [4]	-	-	1207	1222		
SCface [16]	50	300	80	480		24×24, 32×32, 48×48

Table 2: Rank-1 identification accuracies (%) obtained using Verilook (COTS-I) for cross resolution face recognition. The target resolution is 96×96. The algorithms which do not support the required magnification factor are presented as $'-'$.

	Probe Resolution	Original Image	Bicubic Interp.	Dong *et al.* [9]	Kim *et al.* [20]	Gu *et al.* [18]	Dong *et al.* [8]	Peleg *et al.* [27]	Yang *et al.* [38]	Proposed SHSR
CMU MultiPIE	8×8	0.0±0.0	0.1±0.0	-	-	-	-	-	-	**82.6±1.5**
	16×16	0.0±0.0	1.1±0.1	-	-	-	-	-	-	**91.1±1.3**
	24×24	1.2±0.4	3.1±0.6	2.0±3.5	4.1±1.0	4.3±1.0	4.2±0.6	-	-	**91.8±1.8**
	32×32	3.4±0.6	16.9±1.3	9.7±1.1	17.5±1.1	15.4±1.1	6.9±0.2	8.3±0.9	-	**91.9±1.7**
	48×48	91.9±1.1	95.8±0.4	85.8±0.7	**96.2±0.6**	93.1±0.9	95.5±1.1	92.8±0.4	94.0±0.6	91.5±1.5
CAS-PEAL	8×8	0.0±0.0	0.0±0.0	-	-	-	-	-	-	**92.8±0.7**
	16×16	0.0±0.0	0.2±0.3	-	-	-	-	-	-	**94.4±1.1**
	24×24	0.4±0.6	14.9±1.7	0.4±0.2	2.3±0.8	1.9±0.7	2.5±0.7	-	-	**95.3±1.4**
	32×32	3.7±0.7	76.5±1.8	5.4±1.2	11.8±1.1	8.1±2.3	2.1±0.7	3.1±1.6	-	**95.6±1.1**
	48×48	63.4±1.7	90.8±1.5	46.5±2.5	75.8±2.3	77.7±2.1	72.0±0.7	74.0±2.6	73.3±3.3	**95.4±1.5**
SCface	24×24	1.1±0.2	0.8±0.1	0.4±0.2	0.4±0.2	1.5±0.3	1.3±0.3	-	-	**14.7±3.3**
	32×32	1.8±0.5	2.5±0.3	2.2±0.4	2.0±0.0	2.3±0.3	0.7±0.3	1.8±0.5	-	**15.6±1.3**
	48×48	6.5±0.6	9.5±1.9	6.9±0.6	6.7±1.2	7.7±0.6	7.5±1.3	7.3±0.9	6.8±0.7	**18.5±2.6**

Using Eq. 9 and Eq. 11, and the second level representation for the given image in the target domain, a *synthesized* output of the given image is obtained. First $\boldsymbol{\alpha}_h^{l,test}$ is calculated with the help of $\mathbf{G}_h^2$, and then $\mathbf{x}_h^{test}$ is obtained using $\mathbf{G}_h^l$, which is the synthesized image in the target domain:

$$\boldsymbol{\alpha}_h^{l,test} = \mathbf{G}_h^{\mathbf{2}} \boldsymbol{\alpha}_h^{2,test}; \; \mathbf{x}_h^{test} = \mathbf{G}_h^{\mathbf{1}} \boldsymbol{\alpha}_h^{l,test} \tag{15}$$

It is important to note that the synthesized high resolution image is a sparse combination of the basis functions of the learned high resolution dictionary. In order to obtain a good quality, identity-preserving high resolution synthesis, the dictionary is trained with the pre-acquired high resolution database. As will be demonstrated via experiments as well, a key highlight of this algorithm is to learn good quality, representative dictionaries with a single sample per subject. The high resolution synthesized output image $\mathbf{x}_h^{test}$ can then be used by any face identification engine for recognition.

3. Datasets and Experimental Protocol

The effectiveness of the proposed SHSR algorithm is demonstrated by evaluating the face recognition performance with original and synthesized images. Two commercial-off-the-shelf face recognition systems (COTS), Verilook (COTS-I) [2] and Luxand (COTS-II) [1] are used on four different face databases. For Verilook, the face quality and confidence thresholds are set to minimum, in order to reduce enrollment errors. The performance of the proposed algorithm is compared with six recently proposed super-resolution and synthesis techniques by Kim *et al.* [20][1] (kernel ridge regression), Peleg *et al.* [27][2] (sparse representation based statistical prediction model), Gu *et al.* [18][3] (convolutional sparse coding), Yang *et al.* [38][4] (dictionary learning), Dong *et al.* [8][5] (deep convolutional networks), and Dong *et al.* [9][6] (deep convolutional networks) along with one of the most popular techniques, bicubic interpolation. The results of the existing super-resolution algorithms are computed by using the models provided by the authors at the links provided in the footnotes. It is to be noted that not all the algorithms support all the levels of magnification. For instance, the algorithm proposed by Kim *et al.* [20] supports up to 4 levels of magnification whereas, Yang *et al.*'s algorithm [38] supports up to 2 levels of magnification.

Face Datasets: Table 1 summarizes the statistics of the datasets in terms of training and testing partitions, along with the resolutions. Details of each are provided below:

1. CMU Multi-PIE Dataset [17]: Images pertaining to 337 subjects are selected with frontal pose, uniform illumi-

[1] https://people.mpi-inf.mpg.de/ kkim/supres/supres.htm
[2] http://www.cs.technion.ac.il/ẽlad/software/
[3] http://www4.comp.polyu.edu.hk/ cslzhang/
[4] http://www.ifp.illinois.edu/j̃yang29/
[5] http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html
[6] http://mmlab.ie.cuhk.edu.hk/projects/FSRCNN.html

nation, and neutral expression. 100 subjects are used for training while the remaining 237 are in the test set.

2. CAS-PEAL Dataset [13] consists of face images of 1040 subjects. All subjects have a single, high-resolution *normal* image, and images of different covariates such as lighting, expression, and distance. For this research, *normal* images are used as the high resolution gallery database while face images under the *distance* covariate are downsampled and used as probe images.

3. SCface Dataset [16]: It consists of 130 subjects, each having one high resolution frontal face image and multiple low resolution images, captured from three distances using surveillance cameras.

4. Real World Scenarios Dataset [4] contains images of seven subjects associated with the London Bombing, Boston Bombing, and Mumbai Attacks. Each subject has one high resolution gallery image and multiple low resolution test images. The test images are captured from surveillance cameras, and are collected from multiple sources from the Internet. Since the number of subjects are just seven, in order to mimic a real world scenario, the gallery size is increased to create an *extended* gallery of 1200 subjects. Images from the CMU Multi-PIE, ND Human Identification Set-B [10], and MEDS[11] datasets are used for the same.

Protocol: For all the datasets, a real world matching protocol is followed. For each subject, multiple low resolution images are used as probe images, which are matched against the pre-acquired database of high resolution gallery images. Only a single high resolution image per subject is used as gallery. The proposed and comparative algorithms are used to synthesize (or super-resolve) a high resolution image from a given low resolution input. For all datasets except the SCface dataset, test images are of sizes varying from 8×8 to 48×48. The magnification factor varies from 2 (for probes of 48×48) to 12 (for probes of 8×8) to match it against the gallery database of size 96×96. For the SCface database, probe resolutions are 24×24, 32×32, and 48×48, one corresponding to each distance. Face detection is performed using face co-ordinates (if provided) or using Viola Jones Face Detector [33] and synthetic downsampling is performed to obtain lower resolutions. All the experiments are performed with five times random sub-sampling to ensure consistency.

Implementation Details: The SHSR algorithm is trained using the pre-acquired gallery database for each dataset. The regularization constant for sparsity is kept at 0.85. Different dictionaries have different dimensions, based on the input data. For instance, the two-level dictionaries created for SCface dataset contain 100 and 80 atoms in the first and second dictionary, respectively. The source code of the algorithm will be made publicly available in order to ensure reproducibility of the proposed approach.

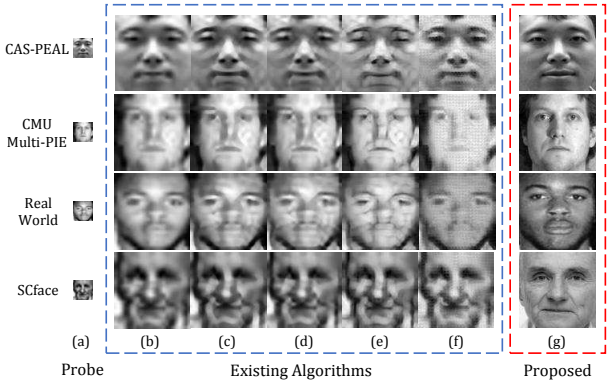

Figure 3: Probe images of 24×24 are super-resolved/synthesized to 96×96. (a) corresponds to the original probe, (b)-(g) correspond to different techniques: bicubic interpolation, Kim *et al* [20], Gu *et al.* [18], Dong *et al.* [8], Dong *et al.* [9], and the proposed SHSR algorithm.

4. Results and Analysis

The proposed algorithm is evaluated with three sets of experiments: (i) face recognition performance with resolution variations, (ii) image quality measure, and (iii) face identification analysis with different dictionary levels. The resolution of the gallery is set to 96×96. For the first experiment, the probe resolution varies from 8×8 to 48×48, while it is fixed to 24×24 for the next two experiments.

4.1. Face Recognition across Resolutions

For all datasets and resolutions, results are tabulated in Tables 2 to 4. Figure 3 shows sample synthesized/super-resolved images from multiple datasets obtained with the proposed and existing algorithms. The key observations pertaining to these set of experiments are presented below:

8×8 and 16×16 probe resolutions: Except bicubic interpolation, none of the super-resolution or synthesis algorithms used in this comparison support a magnification factor of 12 (for 8×8) or 6 (for 16×16); therefore, the results on these two resolutions are compared with original resolution (when the probe is used as input to COTS as it is, without any resolution enhancement) and bicubic interpolation only. As shown in the third and fourth columns of the two tables, on the CMU Multi-PIE and CAS-PEAL datasets, matching with original and bicubic interpolated images results in an accuracy of $\leq 1.1\%$ whereas, the images synthesized using the proposed algorithm provide rank-1 accuracy of 82.6% and 92.8%, respectively (Table 2).

24×24 and 32×32 probe resolutions: As shown in Table 2, on CMU Multi-PIE and CAS-PEAL datasets with test resolution of 24×24 and 32×32, the synthesized images obtained using the proposed SHSR algorithm yield a rank-1 accuracy greater than 91.8%. Other approaches yield a rank-1 accuracy less than 20%, except bicubic interpolation on 32×32 size which provides a rank-1 accuracy of 76.5%.

Table 3: Rank-1 identification accuracies (%) obtained using Luxand (COTS-II) for cross resolution face recognition. The target resolution is 96×96. The algorithms which do not support the required magnification factor are presented as '$-$'.

	Probe Resolution	Original Image	Bicubic Interp.	Dong *et al.* [9]	Kim *et al.* [20]	Gu *et al.* [18]	Dong *et al.* [8]	Peleg *et al.* [27]	Yang *et al.* [38]	Proposed SHSR
CMU Multi-PIE	8×8	0.0±0.0	0.0±0.0	-	-	-	-	-	-	**82.3±1.4**
	16×16	0.0±0.0	0.0±0.0	-	-	-	-	-	-	**90.5±1.1**
	24×24	0.9±0.3	1.0±0.3	2.3±0.5	5.9±0.7	6.8±1.4	6.8±0.5	-	-	**92.1±1.5**
	32×32	11.3±1.1	18.3±7.1	13.5±0.6	28.6±1.2	24.3±2.3	19.4±1.5	17.4±2.5	-	**92.2±1.6**
	48×48	90.2±0.5	**97.9±0.5**	96.0±0.6	97.1±0.7	96.6±0.5	96.9±0.6	97.5±0.5	96.2±0.4	91.9±1.6
CAS-PEAL	8×8	0.0±0.0	0.0±0.0	-	-	-	-	-	-	**91.7±0.9**
	16×16	0.0±0.0	0.6±0.6	-	-	-	-	-	-	**93.3±0.7**
	24×24	0.5±0.4	49.3±1.3	2.3±0.8	10.2±1.1	7.3±1.5	5.8±0.4	-	-	**93.7±1.4**
	32×32	11.1±2.8	92.5±2.0	27.9±1.1	34.6±2.3	31.7±2.3	15.2±3.3	28.0±1.9	-	**93.9±1.3**
	48×48	88.1±0.6	**95.4±1.4**	85.7±2.8	93.3±1.4	93.3±1.4	91.4±1.4	93.6±1.9	90.8±1.7	93.9±1.5
SCface	24×24	1.1±0.2	1.5±1.0	1.2±0.5	0.8±0.1	2.2±0.5	1.9±0.6	-	-	**14.7±2.4**
	32×32	2.2±0.4	3.2±0.4	3.5±0.5	2.6±0.7	4.0±0.7	2.6±1.0	2.8±0.7	-	**15.7±1.3**
	48×48	9.7±1.7	12.6±1.7	9.6±1.1	11.6±1.3	10.1±1.8	11.7±2.0	11.4±1.2	11.9±1.0	**19.1±3.4**

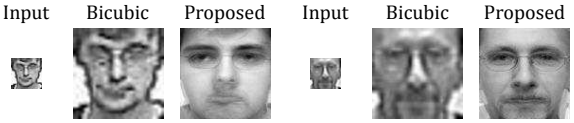

Figure 4: Sample images from SCface dataset incorrectly synthesized to 96×96 by SHSR algorithm for 32×32 input.

As shown in Table 3, similar performance trends are observed using COTS-II on the two databases. For SCface, the rank-1 accuracy with SHSR is significantly higher than the existing approaches; however, due to the challenging nature of the database, both commercial matchers provide low rank-1 accuracies. Figure 4 presents sample images from the SCface dataset, incorrectly synthesized via the proposed SHSR algorithm. Varying acquisition devices of the training and testing partitions, along with the covariates of pose and illumination creates the problem further challenging.

48×48 probe resolution: Using COTS-I, the proposed algorithm achieves improved performance than other techniques, except on the CMU Multi-PIE dataset, where it does not perform as well. On all other databases, the proposed algorithm yields the best results. Upon analyzing both the Tables, it is clear that the proposed algorithm is robust to different recognition systems, and performs well without any bias for a specific kind of recognition algorithm.

Another observation is that with COTS-II, images super-resolved using bicubic interpolation yield best results on the first two databases. However, it should be noted that these results are only observed for a magnification factor of 2 and for images which were synthetically down-sampled. In real world surveillance datasets, such as SCface, the proposed approach performs best with both commercial systems.

Real World Scenarios Dataset: Table 4 summarizes the results of COTS-I on Real World Scenarios dataset. Since the gallery contains images from 1200 subjects, we summarize the results in terms of the identification performance with top 20% retrieved matches. It is interesting to observe that for all test resolutions, the proposed algorithm significantly outperforms existing approaches. SHSR achieves an identification accuracy of 53.3% on probe resolution of 8×8 and an accuracy of 60.0% for 48×48 test resolution.

Cross Dataset Experiments: The SHSR algorithm was trained on the CMU Multi-PIE dataset and tested on the SCface dataset for a probe resolution of 24×24. A rank-1 identification accuracy of 1.62% (1.92%) was obtained with COTS-I (COTS-II), and a rank-5 identification accuracy of 7.54% and 9.06% was obtained, respectively. These results show that the proposed model is able to achieve better recognition performance as compared to other techniques. The drop in accuracy strengthens our hypothesis that using an identity-aware model for image synthesis is more beneficial for achieving higher recognition performance.

4.2. Quality Analysis

Figure 3 shows examples of synthesized/super-resolved images from multiple datasets generated using the proposed and existing algorithms. In this figure, images of 96×96 are synthesized from low resolution images of 24×24. It is observed that the images obtained using existing algorithms (columns (b) - (f)) have artifacts in terms of blockiness and/or blurriness. However, the quality of images obtained using the proposed algorithm (column (g)) are significantly better. To compare the visual quality of the outputs, a no reference image quality measure, BRISQUE [24] is utilized. Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) computes the distortion in the image by using the statistics of locally normalized luminance coefficients. It is calculated in the spatial domain and is used to estimate the loss of naturalness in an image. A lower value denotes less distortions in the image. From Table 5, it can be seen that images obtained using the proposed SHSR algorithm have a better (lower) BRISQUE score as compared

Table 4: Real World Scenarios: Recognition accuracy obtained in top 20% retrieved matches, against a gallery of 1200 subjects using COTS-I (Verilook). The gallery database has resolution of 96×96.

Probe Resolution	Original Image	Bicubic Interpolation	Dong *et al.* [9]	Kim *et al.* [20]	Gu *et al.* [18]	Dong *et al.* [8]	Peleg *et al.* [27]	Yang *et al.* [38]	Proposed SHSR
8×8	0.0	0.0	-	-	-	-	-	-	**53.3**
16×16	0.0	13.3	-	-	-	-	-	-	**53.3**
24×24	6.6	16.6	13.3	26.6	13.3	6.6	-	-	**53.3**
32×32	33.3	16.6	33.3	33.3	33.3	16.6	40.0	-	**53.3**
48×48	33.3	46.6	26.6	33.3	26.6	20.0	33.3	40.0	**60.0**

Table 5: Average no reference quality measure - BRISQUE [24] for probe resolution of 24×24 synthesized to 96×96, obtained over five folds. A lower value for BRISQUE corresponds to lesser distortions in the image.

Dataset	Bicubic Interp.	Dong *et al.* [9]	Kim *et al.* [20]	Gu *et al.* [18]	Dong *et al.* [8]	Proposed SHSR
CMU Multi-PIE	54.8 ± 0.1	28.94 ± 0.0	50.8 ± 0.1	52.8 ± 0.1	48.8 ± 0.1	$\mathbf{26.2 \pm 1.3}$
CAS-PEAL	60.0 ± 0.2	52.86 ± 0.0	54.3 ± 0.2	56.4 ± 0.1	53.4 ± 0.1	$\mathbf{39.3 \pm 0.3}$
SCface	58.7 ± 0.1	52.86 ± 0.0	53.2 ± 0.2	54.9 ± 0.1	47.2 ± 0.1	$\mathbf{34.2 \pm 0.6}$
Real World	57.5	28.94	54.5	54.6	49.54	**25.9**

Table 6: Rank-1 accuracies (%) for varying levels of SHSR algorithm with 24×24 probe and 96×96 gallery.

Dataset	COTS	Dictionary Levels		
		$k = 1$	$k = 2$	$k = 3$
CMU Multi-PIE	Verilook	91.4	**91.8**	**91.8**
	Luxand	92.0	92.1	**92.5**
CAS-PEAL	Verilook	93.8	**95.3**	93.7
	Luxand	92.2	**93.7**	93.6
SCface	Verilook	15.0	14.7	**15.2**
	Luxand	**15.6**	14.7	15.3

to images generated with existing algorithms.

4.3. Effect of Dictionary Levels

As explained in the algorithm section, synthesis can be performed at different levels of hierarchical dictionary, i.e. with varying values of k. This experiment is performed to analyze the effect of different dictionary levels on identification performance. The proposed algorithm is used to synthesize high resolution images (96×96, magnification factor of 4) from input images of size 24×24 with varying dictionary levels, i.e. $k = 1, 2, 3$. First level dictionary ($k = 1$) is equivalent to shallow dictionary learning, whereas two and three levels correspond to synthesis with hierarchical dictionary learning. Table 6 reports the rank-1 identification accuracies obtained with the two commercial matchers for three datasets. The results show that the proposed approach with $k = 2$ generally yields the best results. In some cases, the proposed approach with $k = 3$ yields better results. However, computational complexity with 3-level hierarchical dictionary features is higher and the improvement in accuracy is not consistent across datasets. On the other hand, paired t-test on the results obtained by the shallow dictionary and 2-level hierarchical dictionary demonstrate statistical significance with a confidence level of 95% (for Verilook). Specifically, for a single image, synthesis with level-1 dictionary requires 0.42 ms, level-2 requires 0.43 ms, and level-3 requires 0.45 ms.

5. Conclusion and Future Research

The key contribution of this research is a recognition-oriented pre-processing module based on dictionary learning algorithm for synthesizing a high resolution face image from a low resolution input. The proposed SHSR algorithm learns the representations of low and high resolution images in a hierarchical manner, along with a transformation between the deepest representations. The results are demonstrated on four datasets with test image resolution ranging from 8×8 to 48×48. Matching these requires generating synthesized high resolution images with a magnification factor of 2 to 12 for gallery images of dimension 96×96. Results computed in terms of face recognition performance and image quality measure illustrate that the proposed algorithm consistently yields good recognition results. Computationally, the proposed algorithm requires less than 1 millisecond for generating a synthesized high resolution image which further showcases the efficacy and usability of the algorithm for low resolution face recognition applications. In future, we plan to extend the proposed synthesis based approach for (i) face recogntion in videos for frame selection and enhancement [15], (ii) disguised face recognition [6, 7, 21] where it can also be used to remove the effect of disguise, and (iii) face recognition in low resolution near-infrared images [14].

6. Acknowledgment

This research is partially supported by MEITY (Government of India). M. Vatsa and R. Singh are partially supported through Infosys Center for Artificial Intelligence, IIIT-Delhi. S. Nagpal is supported by TCS PhD Fellowship.

References

[1] Luxand. https://www.luxand.com. 5

[2] Verilook. http://www.neurotechnology.com/verilook.html. 5

[3] S. Baker and T. Kanade. Hallucinating faces. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000. 2

[4] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha. Improving cross-resolution face matching using ensemble-based co-transfer learning. *IEEE Transactions on Image Processing*, 23(12):5654–5669, 2014. 1, 2, 5, 6

[5] R. Dahl, M. Norouzi, and J. Shlens. Pixel recursive super resolution. In *IEEE International Conference on Computer Vision*, 2017. 2

[6] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa. Disguise detection and face recognition in visible and thermal spectrums. In *International Conference on Biometrics*, 2013. 8

[7] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLOS ONE*, 9, 07 2014. 8

[8] 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*, 38(2):295–307, 2016. 2, 5, 6, 7, 8

[9] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*, pages 391–407, 2016. 2, 5, 6, 7, 8

[10] P. J. Flynn, K. W. Bowyer, and P. J. Phillips. Assessment of time dependency in face recognition: An initial study. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 44–51, 2003. 6

[11] A. Founds, N. Orlans, G. Whiddon, and C. Watson. NIST special database 32-multiple encounter dataset II (MEDSII). *National Institute of Standards and Technology, Tech. Rep*, 2011. 6

[12] T. C. Fu, W. C. Chiu, and Y. C. F. Wang. Learning guided convolutional neural networks for cross-resolution face recognition. In *IEEE International Workshop on Machine Learning for Signal Processing*, 2017. 2

[13] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(1):149–161, 2008. 5, 6

[14] S. Ghosh, R. Keshari, R. Singh, and M. Vatsa. Face identification from low resolution near-infrared images. In *IEEE International Conference on Image Processing*, pages 938–942, 2016. 8

[15] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa. MDLFace: Memorability augmented deep learning for video face recognition. In *IEEE International Joint Conference on Biometrics*, 2014. 8

[16] M. Grgic, K. Delac, and S. Grgic. SCface - surveillance cameras face database. *Multimedia Tools Application*, 51(3):863–879, 2011. 5, 6

[17] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image Vision Computing*, 28(5):807–813, 2010. 5

[18] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang. Convolutional sparse coding for image super-resolution. In *IEEE International Conference on Computer Vision*, 2015. 2, 5, 6, 7, 8

[19] 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*, 25(11):1761–1772, 2015. 2

[20] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, 2010. 5, 6, 7, 8

[21] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa. Disguise faces in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018. 8

[22] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2

[23] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808. 2007. 4

[24] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 7, 8

[25] S. P. Mudunuri and S. Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(5):1034–1040, 2016. 2

[26] J. Ngiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. Y. Ng. Sparse filtering. In *Advances in Neural Information Processing Systems*, pages 1125–1133. 2011. 2

[27] T. Peleg and M. Elad. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Transactions on Image Processing*, 23(6):2569–2582, 2014. 2, 5, 7, 8

[28] G. Polatkan, M. Zhou, L. Carin, D. Blei, and I. Daubechies. A bayesian nonparametric approach to image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):346–358, 2015. 2

[29] R. Rubinstein, M. Zibulevsky, and M. Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 58(3):1553–1564, 2010. 2, 3

[30] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa. Deep dictionary learning. *IEEE Access*, 4:10096–10109, 2016. 2

[31] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias. Multilevel dictionary learning for sparse representation of images. In *Digital Signal Processing and Signal Processing Education Meeting*, pages 271–276, 2011. 2

[32] T. Tong, G. Li, X. Liu, and Q. Gao. Image super-resolution using dense skip connections. In *IEEE International Conference on Computer Vision*, 2017. 2

[33] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal Computer Vision*, 57(2):137–154, 2004. 6

[34] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. A comprehensive survey to face hallucination. *International Journal of Computer Vision*, 106(1):9–30, 2014. 2

[35] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2216–2223, 2012. 2

[36] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang. Studying very low resolution recognition using deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 1, 2

[37] Z. Wang, Z. Miao, Q. M. Jonathan Wu, Y. Wan, and Z. Tang. Low-resolution face recognition: a review. *The Visual Computer*, 30(4):359–386, 2013. 1, 2

[38] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. 2, 5, 7, 8

[39] 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 *International Conference on Biometrics*, pages 237–242, 2015. 2