Simultaneous Super-Resolution and Feature Extraction for Recognition of Low-Resolution Faces

Pablo H. Hennings-Yeomans
ECE Department
Carnegie Mellon University
phh@cmu.edu

Simon Baker
Microsoft Research
Microsoft Corporation
sbaker@microsoft.com

B.V.K. Vijaya Kumar
ECE Department
Carnegie Mellon University
kumar@ece.cmu.edu

Abstract

Face recognition degrades when faces are of very low resolution since many details about the difference between one person and another can only be captured in images of sufficient resolution. In this work, we propose a new procedure for recognition of low-resolution faces, when there is a high-resolution training set available. Most previous super-resolution approaches are aimed at reconstruction, with recognition only as an after-thought. In contrast, in the proposed method, face features, as they would be extracted for a face recognition algorithm (e.g., eigenfaces, Fisherfaces, etc.), are included in a super-resolution method as prior information. This approach simultaneously provides measures of fit of the super-resolution result, from both reconstruction and recognition perspectives. This is different from the conventional paradigms of matching in a low-resolution domain, or, alternatively, applying a super-resolution algorithm to a low-resolution face and then classifying the super-resolution result. We show, for example, that recognition of faces of as low as 6×6 pixel size is considerably improved compared to matching using a super-resolution reconstruction followed by classification, and to matching with a low-resolution training set.

1. Introduction

In many surveillance scenarios people may be far from the camera and the images of their faces may be small in the field of view. Such low resolution can seriously degrade the performance of conventional face recognition systems. Another similar scenario is when face recognition is used to help automatically organize family photographs, where often faces can be small.

In this paper we study the problem of matching a low-resolution probe image to a high-resolution gallery of enrolled faces. There are two standard approaches to this problem (see Figure 1): (1) Use super-resolution or interpolation to reconstruct a higher resolution version of the low resolution probe and then perform matching in the usual way at higher resolution. (2) Downsample the entire gallery and then perform matching in low resolution.

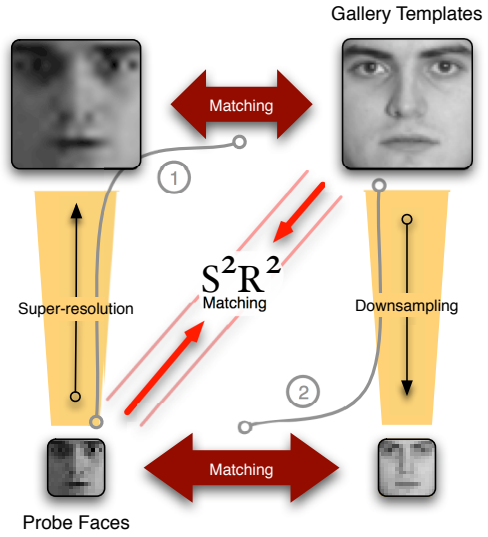

Figure 1. Standard approaches to matching a low resolution probe to a high resolution gallery. (1) Upsampling the probe (interpolation or super-resolution) and then matching. (2) Downsampling the gallery and then matching. In this paper we propose an alternative algorithm that can outperform these two approaches.

Super-resolution methods produce a reconstructed high-resolution image from a low-resolution one [2], or a sequence of images [7, 18], by making assumptions about the image structure or content. An overview of super-resolution is given in [8]. Even though super-resolution reconstruction can be visually appealing, most super-resolution methods are not designed for face recognition, even the ones with face priors, such as in [2, 5, 12]. Recently proposed, the approach in [1], requires a video sequence at enrollment, so is not suited for current face recognition demands [14].

In this paper, we propose a new approach for recogni-

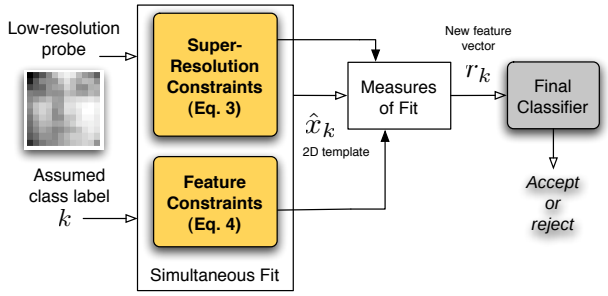

Figure 2. Proposed procedure for recognition of low-resolution face images using standard feature extraction and classification with super-resolution methods.

tion of low-resolution faces. We assume super-resolution and face recognition algorithms have been selected in advance (referred herein as the *base* super-resolution algorithm and the *base* classifier). Our algorithm then uses face features, as would be used by the base classifier, and super-resolution priors of the base super-resolution algorithm, to extract a high-resolution template that simultaneously fits both, super-resolution and face-features constraints. From this template a new set of distortion-based (quality-of-fit) features are computed for recognition (see Fig. 2).

Our approach is fundamentally different from that in [10] because their approach is not simultaneous, but sequential: first, compute high-resolution tensor space parameters from low-resolution ones, and then perform recognition or super-resolution. We do not perform super-resolution in the MSE sense, but use super-resolution models as constraints simultaneously with face-features constraints. Moreover, our approach is more general because it is not limited to a generative multi-factor model, but can also be discriminative.

We show that our algorithm performs far better than first upsampling the image with super-resolution and then matching (Figure 1, path 1). In a database of 224 different faces with illumination variations we obtain 80.7% rank-1 identification accuracy (IDA) while Tikhonov super-resolution [18] gives 54.2%. We also show that the proposed algorithm performs better than matching in low resolution (Figure 1, path 2). In the same example above, using a low-resolution training set yields 72.9% IDA.

One of the key questions underlying our algorithm is whether it is better to match in the low-resolution domain or using the training set at the base resolution (the resolution of the base classifier). This suggest the possibility of using the training set at **both** low and base resolutions. Intermediate resolutions could potentially be used too. We show in Section 4 that our algorithm can facilitate this. Empirical evidence suggests that in some cases matching in multiple resolutions can give lower error rates outperforming even the matching with probe images at base resolution (which in reality are not available).

2. Problem Statement

Face features can be linear or non-linear, however, in this work we use only linear features, such as eigenfaces and Fisherfaces [17, 4]. In our notation we write $\mathbf{F}x_v = f_v$, using the feature matrix $\mathbf{F}$, where v is either g or p for gallery or probe[1]. Thus, f_v is the feature vector that represents x_v. If necessary, a superscript (k) is used to indicate the class.

Face recognition can refer to face verification or face identification. In practice, face verification algorithms can be thought of as computing a distance metric $D(\cdot,\cdot)$ and comparing it to a threshold, τ:

$$D(\mathbf{F}x_g^{(k)}, \mathbf{F}x_p) = D(f_g^{(k)}, f_p) \lessgtr \tau, \qquad (1)$$

where $f_g^{(k)}$ is a feature vector derived from a gallery image that represents class k, and f_p is the feature vector of the input probe face.

In face identification (rank-1), a class-label is produced by computing

$$\arg\min_k D(f_g^{(k)}, f_p) \quad k = 1, \ldots, K \qquad (2)$$

when there are K classes enrolled in the gallery set.

In reality, some times the ideal probe x_p is not available. Instead, we may have a lower resolution version, y_p, and we cannot compute f_p directly using $\mathbf{F}$ and y_p because of the difference in dimensionality between x_p and y_p. Then, the standard procedures illustrated by the two paths in Figure 1 are intuitive algorithms to recognize y_p. We show, however, that the proposed method can be a better alternative.

2.1. Baseline Algorithms

A standard approach to classify y_p is to produce an estimate $\tilde{x}_p$ of the desired x_p from y_p, and apply $\mathbf{F}$ to compute $D(\mathbf{F}x_g, \mathbf{F}\tilde{x}_p)$. This is super-resolution or interpolation followed by classification.

Because of its mathematical simplicity and to show the effectiveness of the proposed approach, we use Tikhonov regularization (see, for example, [18]) as the base super-resolution in the proposed algorithm. Tikhonov regularization obtains $\tilde{x}_p$ by minimizing the objective function

$$\|\mathbf{B}x - y_p\|^2 + \alpha^2\|\mathbf{L}x\|^2, \qquad (3)$$

where matrix $\mathbf{B}$ is the linear model for converting high-resolution image x to its low-resolution version y. Our model takes into account a decoupled *point spread function* (PSF) of the camera as in [3, 2], i.e., a Gaussian kernel is used as the lens PSF and an averaging kernel for the sensor PSF. Here, $\mathbf{L}x$ is a vector of edge values which may be

[1]The gallery is the set containing face images enrolled in a face recognition system, while the probe set contains images of these faces unseen by the system at enrollment, and is used for evaluation.

obtained with first or second derivative approximations, or other kernels, leading to different instances of $\mathbf{L}$. The scalar α is a regularization parameter [18].

Another baseline alternative to classify y_p is to downsample x_g to obtain y_g, with the same resolution as of y_p. This requires downsampling the training images and computing a new feature matrix $\mathbf{F}_L$ (in the low-resolution domain). Then $D(\mathbf{F}_L y_g, \mathbf{F}_L y_p)$ can be computed and recognition is carried out in the usual manner, as illustrated in Figure 1, path 2. Matching with low-resolution faces, however, is undesirable since we assume that x_g and the whole training set are at an optimal resolution for classification, and that reducing the resolution will degrade the recognition performance.

3. Proposed Algorithm: $\mathbf{S^2R^2}$

In this work we propose to classify the low-resolution y_p while using the high-resolution training set ($\mathbf{F}$) in a framework that includes the classifier's metric. In this paper we assume a metric of the form

$$D(\mathbf{F}x, \mathbf{F}x_g^{(k)}) = \|\mathbf{F}x - f_g^{(k)}\|^2, \tag{4}$$

which can be seen as a function of x with parameter k, the claimed or assumed class. Our framework can be extended to use other distance functions, but this is left as future work.

If we look closely at what a classifier is trying to do [6], we know that a good classifier design ideally achieves the following for all the probe and galley images, x_p, x_g:

$$\begin{array}{lll} D(\mathbf{F}x_p, \mathbf{F}x_g) & \text{should be small if} & \Omega(x_p) = \Omega(x_g) \\ D(\mathbf{F}x_p, \mathbf{F}x_g) & \text{should be large if} & \Omega(x_p) \neq \Omega(x_g) \end{array} \tag{5}$$

where $\Omega(u)$ denotes the class of u.

This suggests performing a super-resolution procedure by using Eqs. 3 and 4 jointly in a regularized fashion, and using the training set to find the regularization parameters. Furthermore, a new set of features can be defined using the norms in these equations since we would expect all the norms to be small when $\Omega(y_p) = \Omega(x_g)$, but large otherwise, thus giving cues for recognition. This is the intuition behind the proposed new algorithm, which we call simultaneous super-resolution and recognition ($\mathbf{S^2R^2}$).

3.1. Simultaneous Super-Resolution & Recognition

Face identification and face verification share the first step in the proposed approach as follows. Assume that for a given low-resolution probe face, y_p, the kth class was claimed and we can either compute or look up the features $f_g^{(k)} = \mathbf{F}x_g$ using the gallery. The first step of our algorithm is to find the minimizer, $\hat{x}_p^{(k)}$, of

$$\|\mathbf{B}x - y_p\|^2 + \alpha^2\|\mathbf{L}x\|^2 + \beta^2\|\mathbf{F}x - f_g^{(k)}\|^2, \tag{6}$$

where $\mathbf{B}$, $\mathbf{L}$ and α are defined as in Eq. 3, and β is an additional regularization parameter. The third term in Eq. 6 is a measure of the difference between ideal features for the claimed class and the features that would be produced by the super-resolution result.

Then, with the resulting $\hat{x}_p^{(k)}$, for the case of face verification we can produce a binary decision with

$$w \cdot q(\hat{x}_p^{(k)}) \;\rightarrow\; accept/reject \tag{7}$$

where $q(\hat{x}_p^{(k)})$ is a vector of measures of fit, defined here as the residual norm on each set of *model* assumptions, and can be written as

$$q(\hat{x}_p^{(k)}) = \begin{bmatrix} \|\mathbf{B}\hat{x}_p^{(k)} - y_p\|^2 \\ \|\mathbf{L}\hat{x}_p^{(k)}\|^2 \\ \|\mathbf{F}\hat{x}_p^{(k)} - f_g^{(k)}\|^2 \end{bmatrix}. \tag{8}$$

The dot in Eq. 7 represents an inner-product, and in this work, the vector w is a linear discriminant, but other classifiers may be used in general. Although, $q(\hat{x}_p^{(k)})$ is a reduced feature vector we use for convenience, $\hat{x}_p^{(k)}$ defines a domain where other features can also be extracted. In Eq. 8, the first component measures the fit between the observed low-resolution probe image and the low-resolution version of the resulting super-resolved image. The second component measures the smoothness of the super-resolution result, and the third component measures the difference between the features derived from high-resolution gallery images and those obtained from the super-resolved image.

For the case of identification, we can use Eq. 6 to compute the template $\hat{x}_p^{(k)}$ for each of the K classes in the gallery set. Then, we produce the class-label for y_p by computing

$$\arg\min_k \; w \cdot q(\hat{x}_p^{(k)}) \quad k = 1, \ldots, K. \tag{9}$$

Eqs. 7 and 9 are analogous to Eqs. 1 and 2, respectively.

3.2. Multiple Resolutions

In Section 5 (specifically, Fig. 6), we observe that face images at lower resolutions than that of the base resolution do have some discriminative information. Since we have the training set and gallery available we could, if we wanted, use it at different resolutions. This leads us to question if we can match simultaneously in both, base and probe resolutions, or even using other intermediate resolutions. Our algorithm can be easily adapted to multiple resolutions. We even show in Section 5 that with multiple resolutions it is possible in some cases to outperform matching with training and evaluation sets at the base resolution.

Using two resolutions, $\mathbf{S^2R^2}$ consists of minimizing

$$\|\mathbf{B}x - y_p\|^2 + \alpha^2\|\mathbf{L}x\|^2 + \beta^2\|\mathbf{F}x - f^{(k)}\|^2 + \gamma^2\|\mathbf{F}_L\mathbf{B}x - f_L^{(k)}\|^2 \tag{10}$$

where we have dropped the subscript g, and $\mathbf{F}_L$ is the feature extraction matrix computed with the low-resolution training set, such that $\mathbf{F}_L y_g = \mathbf{F}_L \mathbf{B} x_g = f_L^{(k)}$. Then, with the resulting minimizer, we compute the model norms in Eq. 10 and modify Eq. 8 to define a 4D feature vector accordingly. Recognition involves using a classifier in this new 4D feature space.

In general, for P resolutions, $\mathbf{S^2R^2}$ is implemented with the generalization of Eq. 10 to

$$\|\mathbf{B}x - y_p\|^2 + \alpha^2 \|\mathbf{L}x\|^2 + \sum_{j=1}^{P} \gamma_j^2 \|\mathbf{F}_j \mathbf{D}_j x - f_j^{(k)}\|^2, \quad (11)$$

where again we dropped the subscript g and use the subscript j to indicate the resolution of the training set. For every resolution, $\mathbf{D}_j$ is a formation matrix necessary to compute the corresponding gallery features $f_j^{(k)}$, and $\alpha, \gamma_1, \ldots, \gamma_P$ are regularization parameters. The feature vector q, in this case, will have $P + 2$ residual values.

4. Implementation

To implement the proposed method, outlined in Sections 3.1 and 3.2 and Eqs. 6 to 11, what is needed is: (1) learn the classifier w on the space of predefined residuals, and (2) learn the regularization parameters. We now describe each of these in turn. We also describe how the optimization of equations of the form of Eq. 11 can be computed efficiently.

4.1. Learning the Classifier w

Assume for now that the regularization parameters (α, β, γ) are known. We use a generic training set disjoint from the gallery. For the purposes of learning w, we separate this generic training set into disjoint gallery and probe sets, and we compute a set of $q(\hat{x}_p^{(k)})$ features as described above for all the classes available in this training set. Depending on the recognition task, the equal error rate (EER) or the identification accuracy (IDA) are used to score the performance of a candidate w. In this work, to show the effectiveness of the proposed approach, we define w as a linear discriminant. We select w out of a collection of discriminants as follows.

Consider that, as explained above, $q(\hat{x}_p^{(k)})$ is the vector of residuals that represents y_p when class k was claimed out of K classes, such that for every y_p we can have one true claim and $K - 1$ false claims. For all images in the probe set we can generate in this manner a set of authentic vectors $q(\hat{x}_p^{(k)})$'s and a set of impostor vectors.

We begin with the Fisher linear discriminant of this two-class classification task using all the probe set. And we compute its EER or IDA. We then process each class in turn. Take the $q(\hat{x}_p^{(k)})$'s of probe images, y_p's, from the first class in the probe set, and compute the Fisher discriminant that ideally will separate these $q(\hat{x}_p^{(k)})$'s into true and false claims. And again assign its EER or IDA as a score (using all the probe set to compute this score as well). We repeat this process for the rest of the classes in the probe set. Finally, we also compute the average vector of these probe-class discriminants and score it. Then, the discriminant with the best score is selected. To assign the final w, we search close to the selected discriminant to find a better vector (if possible), since the distributions of $q(\hat{x}_p^{(k)})$ are not necessarily Gaussian and there may be a better discriminant than Fisher's. Figure 3 illustrates this procedure.

4.2. Learning Regularization Parameters

To select the regularization parameters we implemented a parameter search using Powell's enhanced method [15], since it allows us to find minima of functions for which derivatives are not available. We learn the regularization parameters using the same generic training data used in Section 4.1 (divided the same way into probe and gallery). For every vector of parameters sampled by the search, we follow the process of finding w explained above. Once we have the optimal w, its EER or IDA, which we have computed with the generic training set, is saved as a score for this parameter set. Figure 4 shows an example IDA surface sampled with a sparse (2D) grid search for the case of α and β in Eq. 6. With Powell's method we are able to find parameters more efficiently.

4.3. Computational Speed-Ups

Face identification in large databases requires fast computations. Finding the minimum of Eqs. 6, 10 and 11 needs to be computationally efficient because the optimization procedure is performed for every probe, for every available class in the gallery set. Although, gradient descent methods can be relatively fast, they are not fast enough here when the number of comparisons is large. The following singular value decomposition (SVD)-based method results in a faster implementation.

Let's take the case of $\mathbf{S^2R^2}$ with multiple resolutions. Define the composite matrix and composite vector, respectively, as

$$\mathbf{G} = \begin{bmatrix} \mathbf{B} \\ \alpha\mathbf{L} \\ \gamma_1 \mathbf{F}_1 \mathbf{D}_1 \\ \vdots \\ \gamma_P \mathbf{F}_P \mathbf{D}_P \end{bmatrix} \qquad d = \begin{bmatrix} y_p \\ 0 \\ \gamma_1 f_1^{(k)} \\ \vdots \\ \gamma_P f_P^{(k)} \end{bmatrix} \quad (12)$$

such that finding the minimizer of Eq. 11 is equivalent to finding the least-squares solution of $\mathbf{G}x = d$. Using SVD

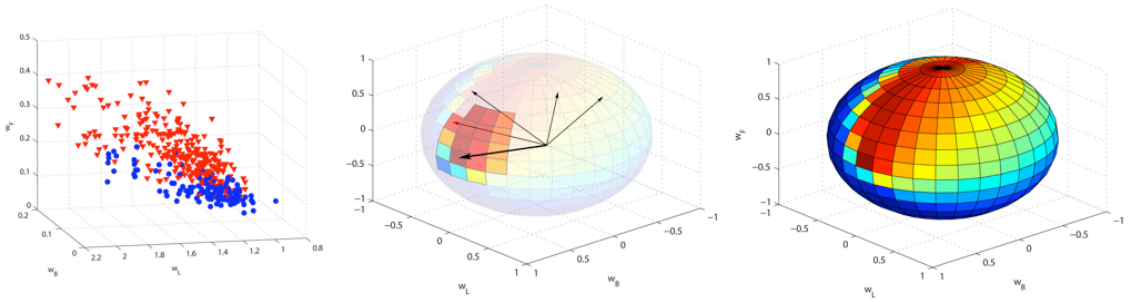

Figure 3. Training the classifier w: For a given α, β pair in Eq. 6, residual vectors $q(\hat{x}_p^{(k)})$ are computed from a training set (*left*) and separated into authentic (blue) and impostors (red) in different manners (see text). For each of these two-class problems, the Fisher discriminant is scored by computing it's IDA (or EER). The discriminant with the best score is selected and a search around it's neighborhood further adjusts the vector to select the best w (*middle*). If the search were conducted exhaustively, in the case of using Eqs. 6 and 8, a sphere's surface will be spanned (*right*). The color at every point in the surface is the training identification accuracy achieved by w, defined as the vector from the origin to that point in the sphere. In this figure, red represents higher accuracy while blue represents lower accuracies. The regularization parameters used are $\alpha = 0.1$ and $\beta = 0.5$.

to write $\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, then

$$\hat{x}_p^{(k)} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^H d = \mathbf{G}^\dagger d. \tag{13}$$

If we split the pseudo-inverse $\mathbf{G}^\dagger$ by groups of columns:

$$\mathbf{G}^\dagger = [\mathbf{G}_y^\dagger \;\; \mathbf{G}_0^\dagger \;\; \mathbf{G}_1^\dagger \;\; \cdots \;\; \mathbf{G}_P^\dagger], \tag{14}$$

then, we can write

$$\hat{x}_p^{(k)} = \mathbf{G}_y^\dagger y_p + m^{(k)} \tag{15}$$

where the class-specific vector $m^{(k)} = \sum_{j=1}^{P} \gamma_j \mathbf{G}_j^\dagger f_j^{(k)}$ can be computed beforehand in the training stage. Thus, for an input probe y_p and one class comparison, computing the template $\hat{x}_p^{(k)}$ requires a single matrix-vector multiplication and one addition. Furthermore, it has complexity $O(p_l p_h)$, p_l and p_h being the number of pixels of the low-resolution probe image y_p and images in the base-resolution training set, respectively; this is regardless of the number of linear constraints in Eq. 11, i.e., regardless of the number of features used in the base classifier.

5. Experiments and Results

The main motivation of this work is low-resolution face recognition. We show how the proposed algorithm compares to the baselines in standard single-frame face recognition. We use the Multi-PIE [9], FERET [13], and FRGC [14] databases and use standard feature extraction algorithms, such as Fisherfaces [4] and Eigenfaces [17]. We also implement features using class-dependence feature analysis (CFA) [11], which have been recently proposed for face recognition. In all the experiments shown here, the only image processing performed on the faces is normalization for zero mean and unit energy, besides alignment of the eyes to a common reference.

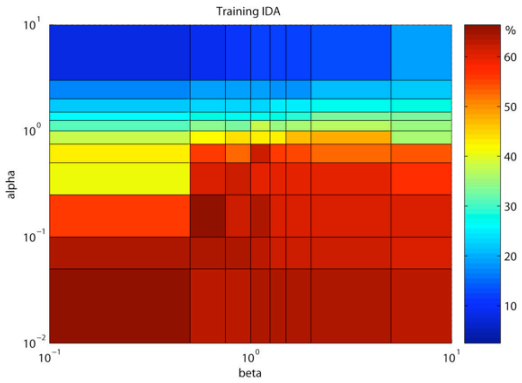

Figure 4. Grid search example for scoring values of α and β in Eq. 6. The color is red for high IDA and blue for low IDA. The value at each point is the one obtained with w (i.e. the one learnt using the procedure in Sec. 4.1, Fig. 3). Instead of using an exhaustive grid search, similar to the one shown here, we implement Powell's enhanced method.

5.1. Multi-PIE

The Multi-PIE database [9] is a recent extension of the PIE database [16]. It has a total of 337 subjects (compared to 68 of PIE) that attended from one to four different recording sessions, each separated by at least a month (unlike PIE, where all images of each subject are captured on the same day in a single session). As in PIE, different face poses, expressions and illumination variations due to flashes from different angles were recorded. Here we use all the subjects available, and we present results using frontal images of neutral expressions with different illuminations. A generic set of 73 subjects is sequestered to compute 25 Fisherfaces for $\mathbf{F}$, and 40 sequestered subjects are used for learning regularization parameters and w. For evaluation (gallery and probe) we use 224 classes, which is the rest of the classes in the database. In our Multi-PIE experiment we have used

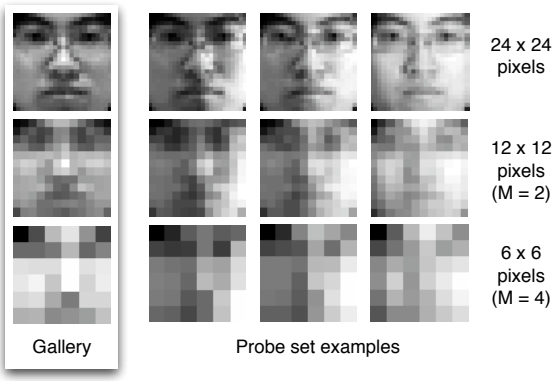

Figure 5. Sample images from the Multi-PIE database [9].

only one image with no flash illumination as gallery, while the probe set contains all the images with flash illumination from all horizontal angles (13 images per subject). Figure 5 is an example for one subject, showing the gallery image and selected probe images at different resolutions. In total, the probe set has 2912 images, which gives 2912 true-class comparisons and 649,376 false-class comparisons.

It is important to note that in this work we assume that in a training set and in the gallery, faces have a resolution that is optimal for classification using $\mathbf{F}$. For example, Figure 6 shows the identification accuracy (IDA) of Fisherfaces using Multi-PIE. It is clear from this result that, after some point, higher resolutions produce practically no improvement for Fisherfaces. Then, considering performance and computational complexity we can define for this case that the optimal resolution of the training set (and the gallery) is 24×24 pixel size.

Using the results of Fig. 6, we downsample the training set to a base resolution of 24×24 pixel size because standard classification performance did not improve considerably at higher resolutions. Thus we define as oracle recognition the ideal case of having the probe faces at this *high* base resolution (shown as HR in the figures of results). This is the resolution in the images of the first row of Fig. 5. We test magnification factors $M = 2$ and $M = 4$ with one and two resolutions in the training set, i.e., using Eq. 6 and Eq. 10, respectively, for the $\mathbf{S^2R^2}$ algorithm.

At all resolutions, we choose $\mathbf{F}$ to represent 25 Fisherfaces (those with the largest eigenvalues).

5.1.1 Basic $\mathbf{S^2R^2}$

In this case the proposed algorithm uses Eq. 6, $M = 2$ with the probe images of size 12×12. Results are shown in Fig. 7. In the left panel we have the IDA. The oracle performance is 87.2% while matching with the training set at resolution of the probe set (shown in the figure under LR) is 72.9%. The proposed algorithm (denoted S2R2s in the figure) achieves an IDA of 80.7%, clearly outperforming LR.

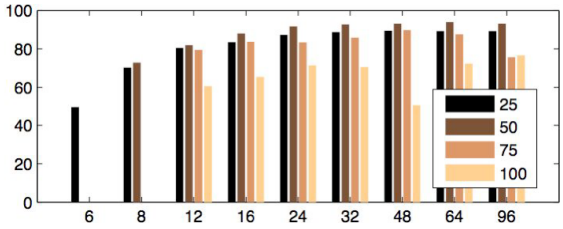

Figure 6. Identification accuracy (IDA) with different number of Fisherfaces. At 6×6 and 8×8, the bars not shown correspond to amounts of Fisherfaces that can't be computed due to the low resolution of the faces [4].

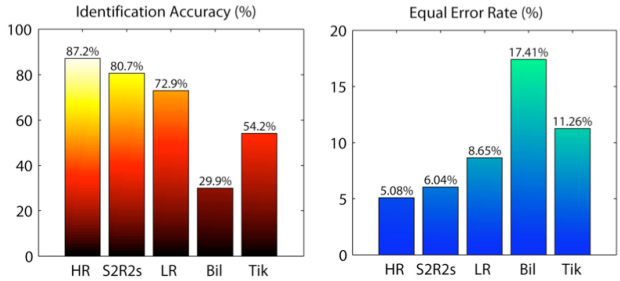

Figure 7. Recognition results with 224 subjects from the Multi-PIE database. (*Left*) Identification accuracy. (*Right*) Equal error rate. From left to right: matching at base-resolution (HR), using the proposed algorithm (Eq. 6, S2R2s), matching at the resolution of the probe (LR), performing bilinear interpolation and then classifying (Bil), and using Tikhonov super-resolution and then classifying (Tik). The probe images are 12×12 pixels.

It is interesting to see how even at small magnification factors both bilinear interpolation, as well as Tikhonov super-resolution perform considerably worse. $\mathbf{S^2R^2}$ outperforms by more than 50% the case of bilinear interpolation and by more than 25% IDA the case of Tikhonov[2].

In the right sub-figure of Fig. 7, the EERs are shown for all algorithms. The trend of performance is similar as in the case of IDA. $\mathbf{S^2R^2}$ is only 0.98% worse in EER than the oracle matching, while the next baseline, matching at low-resolution, is 2.61% worse than $\mathbf{S^2R^2}$.

This results show that the most basic implementation of our algorithm can give better results than standard approaches.

5.1.2 Multiple-Resolutions $\mathbf{S^2R^2}$

In this case, we use both the base-resoluion and low-resolution versions of the gallery, and the multiple resolution algorithm defined by Eq. 11. From Fig. 8, we see that in IDA performance, the proposed algorithm (shown as S2R2m in the figures) is only about 3% below the oracle baseline, while the other algorithms are substantially

[2]We also compared with hallucinating faces [2]. Hallucination performs worse than bilinear interpolation and so the results are omitted.

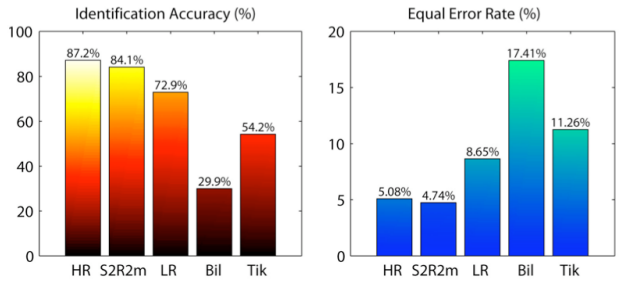

Figure 8. Recognition results with 224 subjects from the Multi-PIE database. (Left panel) IDA. (Right panel) EER. Bar labels as Fig. 7. Here the proposed algorithm using Eq. 10 is shown (S2R2m). Probe images are 12×12 (magnification factor of 2).

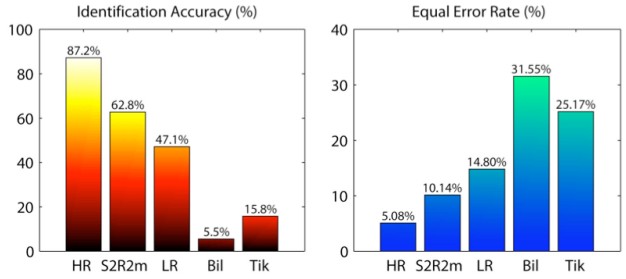

Figure 9. Same as Fig. 8, but for probe images 6×6 ($M = 4$).

worse. But the interesting case is in the EER performance. The proposed algorithm $\mathbf{S^2R^2}$(4.74% EER) has performed better than the oracle matching (5.08% EER). While it may be surprising that $\mathbf{S^2R^2}$ can beat classification in the base resolution, note that $\mathbf{S^2R^2}$ uses Fisherfaces computed at both 24×24 and 12×12 pixels.

Our last result with the Multi-PIE database shows the case of $\mathbf{S^2R^2}$ with multiple resolutions for the harder case of a magnification factor of $M = 4$, i.e. the probe images have 6×6 pixel size. Again, the trend in performance is similar to the case of $M = 2$. While at $M = 4$ the super-resolution algorithms perform far worse than matching in low resolution, $\mathbf{S^2R^2}$ still significantly outperforms matching in low-resolution. One may think that at this point there is no hope for super-resolution algorithms, still our approach can make use of them to produce a better result than matching at low-resolution. In the case of IDA, $\mathbf{S^2R^2}$ obtains 62.8% accuracy while matching in low-resolution gives 47.1%. Furthermore, at rank-3 IDA, i.e., when a list of the top-3 possible identities of the probe are produced, $\mathbf{S^2R^2}$ performs with 81% IDA. Fig. 10, shows this result in an cumulative matching characteristic (CMC) curve.

5.2. FERET

We also evaluated our algorithm with the FERET database [13] FA and FB images, using eigenfaces to show a different feature set. There are 1206 FA images and 1201 FB images for a total of 865 different subjects. We use 400

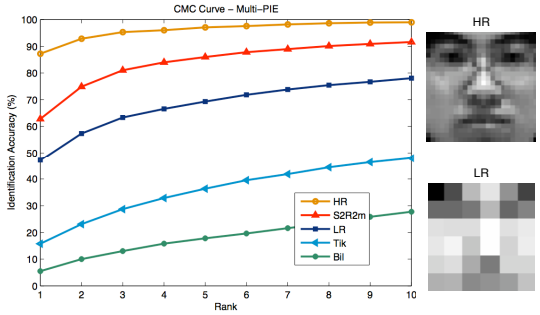

Figure 10. Recognition results with 224 subjects from the Multi-PIE database. Cumulative Matching Characteristic (CMC) curve showing rank-k IDA (percent). Probe images are 6×6 (magnification factor of 4).

classes as a generic set, randomly sampled with replacement, similar to the FERET comptetition. An additional 50 classes were used for the process of finding the regularization parameters. We use 80 eigenfaces at the resolution of the base classifier (HR), and 15 eigenfaces at the resolution of the probes (LR). Figure 11 shows the CMC curve.

We see that the $\mathbf{S^2R^2}$ again outperforms all the baseline algorithms (of course, not counting matching with the probes at the base resolution), although the improvement is not as much as for Fisherfaces relative to LR. This happens in the case of eigenfaces since blurring helps (the already poor features) by smoothing out other distortions, then the search of parameters in $\mathbf{S^2R^2}$ selects a higher α which translates into less distortion-rich templates. Still, $\mathbf{S^2R^2}$ is able to achieve an improvement at such low resolution.

5.3. FRGC Experiment 1

Finally, we also show results obtained with the FRGC database [14], and evaluate $\mathbf{S^2R^2}$ with another set of features. We use a subset of Experiment 1 images with 300 subjects in the gallery and probe sets. Only 1 gallery image is used for each subject, and a total of 5120 probe images, 20 per class if available. We use class-dependence feature analysis (CFA) [11] to extract features, and use 180 features at the resolution of the base classifier (HR), and 30 features at the resolution of the probes (LR). From the database's generic set, 180 classes are used to train CFA, while 42 for learning regularization parameters. Figure 12 shows the CMC curve.

We see that $\mathbf{S^2R^2}$ is able to outperform LR matching by more than 10% rank-1 IDA. The performance of Tikhonov super-resolution and bilinear interpolation is considerably poor. This is because CFA features capture a large amount of detail from faces (high-frequency content) [11], and Tikhonov super-resolution and interpolation produce images that are overly smooth for CFA.

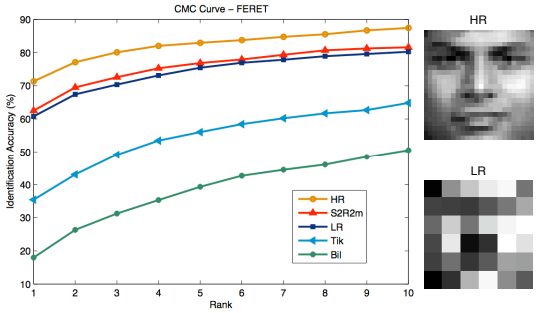

Figure 11. Recognition results with FERET database. Cumulative Matching Characteristic (CMC) curve showing rank-k IDA (percent). Probe images are 6×6 (magnification factor of 4).

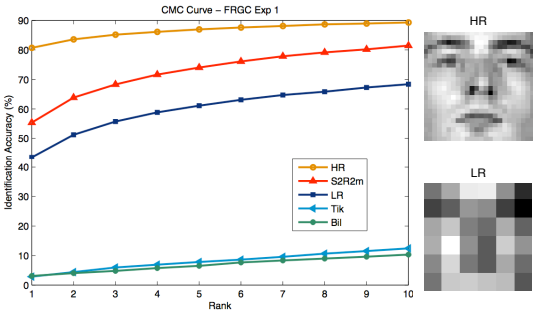

Figure 12. Recognition results with FRGC Experiment 1 database. Cumulative Matching Characteristic (CMC) curve showing rank-k IDA (percent). Probe images are 6×6 (magnification factor of 4).

6. Conclusions

This work effectively gives super-resolution the objective of recognition, rather than just reconstruction. We have proposed an approach for recognition of low-resolution faces that uses super-resolution models together with face recognition features by including them in a regularization framework with the use of the classifier's distance metric. By finding a *high*-resolution template that fits simultaneously into the available models and features under an assumed class membership, we can extract new features for recognition. Our results show that simple linear discriminants using these features produce better recognition performance than standard approaches.

Our formulation can be easily expanded or generalized. It starts from two basic blocks, a super-resolution algorithm and a classification algorithm, and it can include multiple sets of features, and combine super-resolution priors. The application of $\mathbf{S^2R^2}$ to video can be done similarly as in [7] by rewriting Eq. 6 to include multiple input frames. Future work will include an evaluation of $\mathbf{S^2R^2}$ on recognition of faces in video as well as studying sensibility to $\mathbf{B}$.

The extension of $\mathbf{S^2R^2}$ to non-linear features only needs the use of an appropriate non-linear optimization algorithm to minimize Eq. 6. These features may be adapted for specific super-resolution priors.

Acknowledgments

This work is supported in part by CyLab at Carnegie Mellon University.

References

[1] O. Arandjelović and R. Cipolla. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In *ICCV*, Oct. 2007. 1

[2] S. Baker and T. Kanade. Hallucinating faces. In *Fourth Intl. Conf. on Aut. Face and Gesture Rec.*, 2000. 1, 2, 6

[3] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *PAMI*, 24(9):1167 – 1183, 2002. 2

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997. 2, 5, 6

[5] A. Chakrabarti, A. Rajagopalan, and R. Chellappa. Super-resolution of face images using kernel pca-based prior. *IEEE Tran. on Multimedia*, 9(4):888 – 892, June 2007. 1

[6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000. 3

[7] M. Elad and A. Feuer. Super-resolution reconstruction of image sequences. *PAMI*, 21(9):817–834, 1999. 1, 8

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002. 1

[9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Guide to the CMU Multi-PIE database. Technical report, Carnegie Mellon University, 2007. 5, 6

[10] K. Jia and S. Gong. Multi-modal tensor face for simultaneous super-resolution and recognition. In *ICCV*, pages 1683–1690, October 2005. 2

[11] B. Kumar, M. Savvides, and C. Xie. Correlation pattern recognition for face recognition. *Proceedings of the IEEE*, 94(11):1963–1976, Nov. 2006. 5, 7

[12] C. Liu, H.-Y. Shum, and C.-S. Zhang. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. In *CVPR*, volume 1, pages 192–198, 2001. 1

[13] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *PAMI*, 22(10):1090–1104, 2000. 5, 7

[14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, pages 947–954, 2005. 1, 5, 7

[15] M. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.*, 7:155–162, 1964. 4

[16] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database of human faces. Technical report, RI, Carnegie-Mellon, 2001. 5

[17] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991. 2, 5

[18] A. Zomet and S. Peleg. *Super-Resolution Imaging*, chapter "Super-resolution from multiple images having arbitrary mutual motion", pages 195–209. Kluwer, 2001. 1, 2, 3