# Multidimensional Scaling for Matching Low-resolution Face Images

Soma Biswas, *Member, IEEE,* Kevin W. Bowyer, *Fellow, IEEE,*
and Patrick J. Flynn, *Senior Member, IEEE*

## Abstract

Face recognition performance degrades considerably when the input images are of low resolution as is often the case for images taken by surveillance cameras or from a large distance. In this paper, we propose a novel approach for matching low resolution probe images with higher resolution gallery images, which are often available during enrollment, using multidimensional scaling. The ideal scenario is when both the probe and gallery images are of high enough resolution to discriminate across different subjects. The proposed method simultaneously embeds the low resolution probe images and the high resolution gallery images in a common space such that the distances between them in the transformed space approximates the distances had both the images been of high resolution. The two mappings are learned simultaneously from high resolution training images using iterative majorization algorithm. Extensive evaluation of the proposed approach on the Multi-PIE dataset with probe image resolution as low as $8 \times 6$ pixels illustrates the usefulness of the method. We show that the proposed approach improves the matching performance significantly as compared to performing matching in the low-resolution domain or using super-resolution techniques to obtain a higher-resolution test image prior to recognition. Experiments on low resolution surveillance images from Surveillance Cameras Face Database further highlight the effectiveness of the approach.

## Index Terms

face recognition, low-resolution matching, multidimensional scaling, iterative majorization.

S. Biswas, K. W. Bowyer and P. J. Flynn are with the Computer Vision Research Laboratory, Department of Computer Science and Engineering, University of Notre Dame, IN 46556. E-mail: {sbiswas, kwb, flynn}@nd.edu.

# I. INTRODUCTION

Due to its wide range of commercial and law enforcement applications, face recognition has been one of the most important areas of research in the field of computer vision and pattern recognition. Though current algorithms perform well on images captured in controlled environments, their performance is far from satisfactory for images taken under uncontrolled scenarios [1]. Recently, the proliferation of surveillance cameras for security and law-enforcement applications has motivated the development of algorithms which are more suited for handling the kind of images captured by these cameras. Due to the large distance of the camera from the subject, these images usually have very low resolution (LR) face regions which considerably degrades the performance of traditional face recognition algorithms developed for good quality images [2]. Discriminatory properties present in the facial images used for distinguishing one person from the other are often lost due to the decrease in resolution resulting in unsatisfactory performance. On the other hand, high resolution (HR) images of the subjects may be available during enrollment in many cases. Apart from the poor distinguishing properties of the low-resolution images, the resolution difference of the probe and gallery images makes the problem even more challenging. In addition to being able to handle LR images, the algorithms should be efficient enough to handle the large amount of data captured continuously by the surveillance cameras. Though a lot of work has been done on the problem of face recognition to address issues like illumination and pose variations, it is only recently that efforts have been made to deal with matching LR probe images with HR gallery images [3] [2].

There are two standard approaches for addressing this problem: 1) The gallery images are down-sampled to the resolution of the probe images and then recognition is performed. But in this approach, the additional discriminating information available in the high resolution gallery images is lost. 2) The second and more widely used approach relies on super-resolution techniques to obtain higher resolution probe images [4][5] from the low resolution images which are then used for recognition. Most of these techniques aim to obtain a good high resolution reconstruction and are not optimized with respect to recognition performance. Recently there have been efforts to perform super-resolution and recognition simultaneously [2]. But in this approach, given a probe image, the optimization needs to be repeated for each gallery image in the database which might result in high computational overhead, especially for databases of large size.
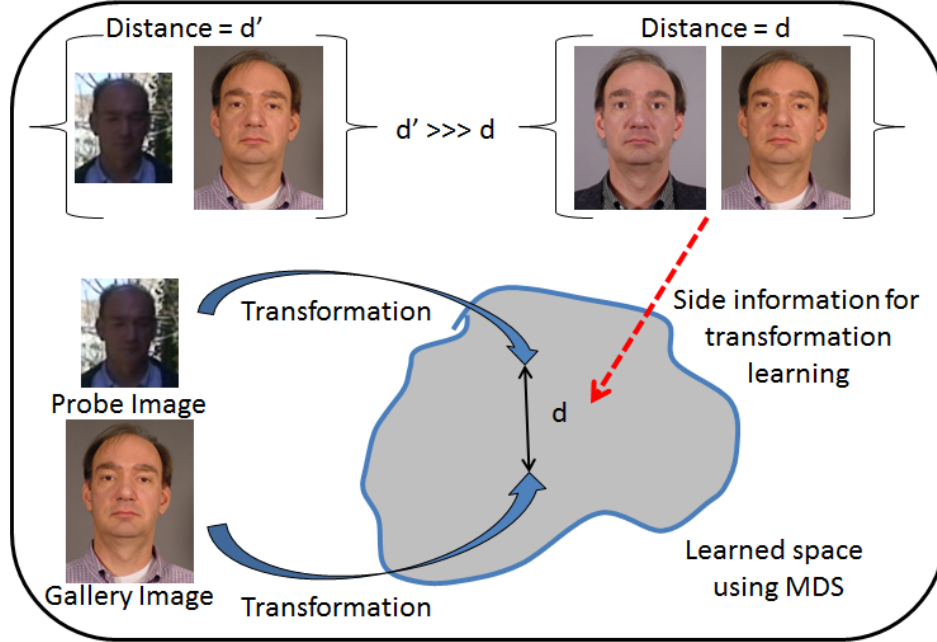
Fig. 1. Overview of the proposed algorithm. The proposed method simultaneously embeds the LR probe and the HR gallery images in a common space such that the distances between them in the transformed space approximate the distances that would have been obtained had both the images been of high resolution.

In this work, we propose a novel approach for matching LR probe images with HR gallery images using Multidimensional Scaling (MDS). The ideal scenario is when both the images are of high resolution thus retaining all the discriminatory properties. The proposed method simultaneously embeds the LR probe and the HR gallery images in a common space such that the distances between them in the transformed space approximates the distances that would have been obtained had both the images been of high resolution (Fig. 1). We show how the two mappings can be learned simultaneously using MDS. The iterative majorization technique is used to learn the transformations from HR training images. The transformations are learned off-line from the training images thus making the approach very efficient. During testing, the input images are transformed using the learned transformation matrix and then matching is performed.

Extensive experiments are performed to evaluate the effectiveness of the proposed approach on the Multi-PIE [6] dataset. The proposed approach improves the matching performance significantly as compared to performing standard matching in the low-resolution domain. The proposed approach also performs considerably better than those of different super-resolution techniques

which obtain higher-resolution images prior to recognition. Performance with different kinds of features shows that the approach is very general and does not depend on the particular kind of input feature used. Comparison with classifier-based approaches further highlights the effectiveness of the approach. Results of experiments on low resolution images from the Surveillance Cameras Face Database [7] are also provided. A preliminary version of this work appeared in [8] which addressed the problem of improving the performance of LR facial images using the distance information between HR images. But that work did not address the problem of matching a LR probe image with HR images in the gallery.

## II. PREVIOUS WORK

In this section, we discuss the related approaches in the literature. Most of the current approaches which address the problem of face recognition from images of low resolution follow a super-resolution approach. These approaches aim to obtain an HR image from the LR input which is then used for recognition. Many of the current super-resolution techniques use face priors to obtain better image reconstruction. Baker and Kanade [4] [9] propose an algorithm to learn a prior on the spatial distribution of the image gradients for frontal images of faces. The gradient prior is learned using a collection of high resolution training images of human faces. Chakrabarti *et al.* [5] propose a learning-based method using kernel principal component analysis for deriving prior knowledge about the face class for performing super-resolution. Liu *et al.* [10] propose a two-step statistical modeling approach for hallucinating an HR face image from an LR input. The relationship between the HR images and their corresponding down-sampled and smoothed LR images is learned using a global linear model and the residual high-frequency content is modeled by a patch-based non-parametric Markov network.

Freeman *et al.* [11] explore training-based super-resolution algorithms in which the fine details corresponding to different image regions seen at a low resolution are learned from a training set and then the learned relationships are used to predict fine details in other images. Chang *et al.* [12] use manifold learning approaches for recovering the HR image from a single LR input. The basic idea is that small image patches in the low and high-resolution images form manifolds with similar local geometry in two distinct feature spaces. The high-resolution embedding is estimated using training image pairs. A novel patch-based face hallucination framework is proposed by Liu *et al.* [13]. First, the relation of the different constituent factors including identity and patch-

locations is learned using a TensorPatch model. Assuming that the low-resolution space and high-resolution space share similar local distribution structure, the estimated parameters are used for synthesizing high-resolution images. Yang *et al.* [14] address the problem of generating a super-resolution image from a low-resolution input image from the perspective of compressed sensing. Regression based techniques like kernel ridge regression have also been used to learn a mapping from input LR images to target HR images from example image pairs [15]. Since many face recognition systems use an initial dimensionality reduction method, Gunturk *et al.* [16] propose eigenface-domain super-resolution in the lower dimensional face space. Some other approaches like support vector data description [17] and advanced correlation filters [18] have also been used to address this problem. Several super-resolution approaches also try to reconstruct a high resolution image from a sequence of input images [19].

The main aim of these techniques is to produce an HR image from the LR input using assumptions about the image content, and they are usually not designed (or optimized) from a matching perspective. Recently, Hennings-Yeomans *et al.* [2] proposed an approach to perform super-resolution and recognition simultaneously. Using features from the face and super-resolution priors, the authors aim to extract an HR template that simultaneously fits the super-resolution as well as the face-feature constraints. Arandjelovic and Cipolla [20] propose a generative model for separating the illumination and down-sampling effects for the problem of matching a face in an LR query video sequence against a set of HR gallery sequences. The illumination model is learned from training data of different individuals with varying illumination while the LR artifact model is estimated on a person-specific basis using enrolled person's training sequences. Given an LR face image, Jia and Gong [3] propose directly computing a maximum likelihood identity parameter vector in the HR tensor space which can be used for recognition and reconstruction of HR face images.

## III. PROBLEM FORMULATION

In this section, we describe the proposed formulation for matching LR probe images with HR images present in the gallery using a MDS approach. We want to transform the HR and LR input images in such a way that the distances between them approximates the distances that would have been obtained had both the images been of high resolution.

We learn the transformations using a training set consisting of HR images and LR images

of the same subjects. The HR and LR images of the same subject need not be obtained under the same imaging conditions. We do not assume any subject overlap between the training set and the gallery set. Let the HR training images be denoted by $\mathbf{I}_i^h$, $i = 1, 2, \ldots \cdots, N$ and the corresponding LR images of the same subjects be denoted by $\mathbf{I}_i^l$, where $N$ is the number of HR or LR images. In the absence of HR and LR images of the same subject, the LR images can be generated from the HR images by downsampling and smoothing. Let $\mathbf{x}^h$ and $\mathbf{x}^l$ denote the features of the HR and LR images respectively. Since the two images to be matched have different resolution, the size of the images as well as the features extracted from the images can have different dimensions and semantics. Given training data, our goal is to find transformations such that the distances between the transformed LR and HR feature vectors $\mathbf{x}_i^l$ and $\mathbf{x}_j^h$ approximates $d_{i,j}^h$. Here $d_{i,j}^h$ denotes the distance between the feature vectors corresponding to the HR images and is given by

$$d_{i,j}^h = D(\mathbf{x}_i^h, \mathbf{x}_j^h) \tag{1}$$

Here the distance measure $D$ can be chosen depending upon the kind of feature used to represent the images.

## IV. PROPOSED APPROACH

Let $\mathbf{f} : R^k \rightarrow R^m$ denote the mapping from the $R^k$ input feature space to the embedded Euclidean space $R^m$. Here $m$ is the dimension of the output transformed space and $k$ denotes the input dimension. We consider the mapping $\mathbf{f} = (f_1, f_2, \cdots, f_m)^T$ to be a linear combination of $p$ basis functions of the form

$$f_i(\mathbf{x}; \mathbf{W}) = \sum_{j=1}^{p} w_{ji} \phi_j(\mathbf{x}) \tag{2}$$

where $\phi_j(\mathbf{x}); j = 1, 2, \cdots, p$ can be a linear or non-linear function of the input feature vector $\mathbf{x}$. Here $[\mathbf{W}]_{ij} = w_{ij}$ is the $p \times m$ matrix of the weights to be determined. The mapping defined by (2) can be written in a compact manner as follows

$$\mathbf{f}(\mathbf{x}; \mathbf{W}) = \mathbf{W}^T \phi(\mathbf{x}) \tag{3}$$

In this case, since the two images to be matched are of different resolutions, two different transformations are required to transform the LR and the HR images to the common space. Let

$\mathbf{W}^h$ and $\mathbf{W}^l$ denote the transformation matrices for the HR and LR feature vectors respectively. Let the basis functions for the HR and LR feature vectors be given by

$$\phi^h(\mathbf{x}^h) = [\phi_1^h(\mathbf{x}^h), \phi_2^h(\mathbf{x}^h), \cdots, \phi_{p_h}^h(\mathbf{x}^h)]^T$$

$$\phi^l(\mathbf{x}^l) = [\phi_1^l(\mathbf{x}^l), \phi_2^l(\mathbf{x}^l), \cdots, \phi_{p_l}^l(\mathbf{x}^l)]^T$$

Here $p_h$ and $p_l$ represent the number of basis functions of the HR and LR feature vectors respectively, which can be same or different. Thus the mappings corresponding to the LR and HR images can be written as

$$\mathbf{f}(\mathbf{x}^h; \mathbf{W}^h) = \mathbf{W}^{h^T}\phi^h(\mathbf{x}^h); \qquad \mathbf{f}(\mathbf{x}^l; \mathbf{W}^l) = \mathbf{W}^{l^T}\phi^l(\mathbf{x}^l) \tag{4}$$

The goal is to transform the feature vectors of the HR and LR images such that the distance between the transformed feature vectors approximates the distance $d_{i,j}^h$. So we want to find the matrices $\mathbf{W}^h$ and $\mathbf{W}^l$ which minimizes the following objective function

$$\mathbf{J}_{\mathrm{DP}}(\mathbf{W}^l, \mathbf{W}^h) = \sum_{i=1}^N \sum_{j=1}^N (q_{ij}(\mathbf{W}^l, \mathbf{W}^h) - d_{i,j}^h)^2 \tag{5}$$

Here $N$ is the number of HR (or LR) training images. The term $q_{ij}(\mathbf{W}^l, \mathbf{W}^h)$ denotes the distance between the transformed feature vectors of the $i^{th}$ LR and $j^{th}$ HR training images and is given by

$$q_{ij}(\mathbf{W}^l, \mathbf{W}^h) = |\mathbf{W}^l\phi^l(\mathbf{x}_i^l) - \mathbf{W}^h\phi^h(\mathbf{x}_j^h)| \tag{6}$$

The distance between the corresponding HR images is $d_{i,j}^h$. Note that the distance $q_{ij}(\mathbf{W}^l, \mathbf{W}^h)$ and thus the optimization function depends on both the transformation matrices.

Since our goal here is to improve the matching performance, the objective function in (5) can be modified to include class information of the training data. Thus a class separability term ($\mathbf{J}_{\mathrm{CS}}$) can be added to the distance preservation term ($\mathbf{J}_{\mathrm{DP}}$) resulting in the following objective function

$$\begin{aligned} \mathbf{J}(\mathbf{W}^l, \mathbf{W}^h) &= \lambda \mathbf{J}_{\mathrm{DP}}(\mathbf{W}^l, \mathbf{W}^h) + (1-\lambda)\mathbf{J}_{\mathrm{CS}}(\mathbf{W}^l, \mathbf{W}^h) \\ &= \lambda \sum_{i=1}^N \sum_{j=1}^N (q_{ij}(\mathbf{W}^l, \mathbf{W}^h) - d_{i,j}^h)^2 \\ &\quad + (1-\lambda)\sum_{i=1}^N \sum_{j=1}^N \delta(\omega_i, \omega_j)q_{i,j}^2(\mathbf{W}^l, \mathbf{W}^h) \end{aligned} \tag{7}$$

The first term $\mathbf{J_{DP}}$ is the distance preserving term which ensures that the distance between the transformed feature vectors approximates the distance $d_{i,j}^h$. $\mathbf{J_{CS}}$ is a class separability term to further facilitate discriminability. $\omega_i$ denotes the class label of the $i^{th}$ class. The parameter $\lambda$ controls the relative effect of the distance preserving and the class separability terms on the total optimization. A simple form of the class separability term is given by [21]

$$
\begin{aligned}
\delta(\omega_i, \omega_j) &= 0; & \omega_i \neq \omega_j \\
&= 1; & \omega_i = \omega_j
\end{aligned}
\tag{8}
$$

This specific form of the class separability term penalizes larger distances between data points of the same class, but does not affect data points of different classes. More sophisticated terms may also be used seamlessly with the formulation.

### A. Combined transformation matrix

The goal is to find the two transformation matrices $\mathbf{W}^h$ and $\mathbf{W}^l$ corresponding to the HR and LR feature vectors. In this section, we show how we can solve for the two transformations simultaneously using a combined transformation matrix $\mathbf{W}$. The distance $q_{i,j}(\mathbf{W}^l, \mathbf{W}^h)$ between the transformed LR and HR feature vectors can be written as

$$
\begin{aligned}
q_{i,j}(\mathbf{W}^l, \mathbf{W}^h) &= |\mathbf{W}^{l^T} \phi^l(\mathbf{x}_i^l) - \mathbf{W}^{h^T} \phi^h(\mathbf{x}_j^h)| \\
&= \left| \begin{pmatrix} \mathbf{W}^{l^T} & \mathbf{W}^{h^T} \end{pmatrix} \begin{pmatrix} \phi^l(\mathbf{x}_i^l) \\ -\phi^h(\mathbf{x}_j^h) \end{pmatrix} \right| \\
&= |\mathbf{W}^T(\bar{\phi}^l(\mathbf{x}_i^l) - \bar{\phi}^h(\mathbf{x}_j^h))|
\end{aligned}
\tag{9}
$$

Here the new matrix $\mathbf{W}$ is a concatenation of the two transformation matrices given by $\mathbf{W}^T = \begin{pmatrix} \mathbf{W}^{l^T} & \mathbf{W}^{h^T} \end{pmatrix}$. The new vectors $\bar{\phi}^l(\mathbf{x}^l)$ and $\bar{\phi}^h(\mathbf{x}^h)$ are given by

$$
\bar{\phi}^l(\mathbf{x}^l) = \begin{pmatrix} \phi^l(\mathbf{x}^l) \\ \mathbf{0}_{p_h \times 1} \end{pmatrix}; \qquad \bar{\phi}^h(\mathbf{x}^h) = \begin{pmatrix} \mathbf{0}_{p_l \times 1} \\ \phi^h(\mathbf{x}^h) \end{pmatrix}
\tag{10}
$$

Here $\mathbf{0}_{k \times 1}$ denotes a vector of zeros of length $k \times 1$. Thus the modified vectors $\bar{\phi}^h(\mathbf{x}^h)$ and $\bar{\phi}^l(\mathbf{x}^l)$ are of equal length even though $\phi(\mathbf{x}^h)$ and $\phi(\mathbf{x}^l)$ may be of different lengths. Now the goal is to compute the transformation matrix $\mathbf{W}$ by minimizing the objective function given

by (7). Separating the terms containing the transformation matrix, the objective function can be rewritten as

$$
\begin{aligned}
\mathbf{J}(\mathbf{W}) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i,j}(q_{i,j}(\mathbf{W}) - \beta_{i,j}d_{i,j}^{h})^2 \\
&+ \lambda\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{i,j}(d_{i,j}^{h})^2
\end{aligned}
\tag{11}
$$

The expression for the different terms are given by

$$
\begin{aligned}
\alpha_{i,j} &= (1-\lambda)\delta(\omega_i,\omega_j) + \lambda \\
\beta_{i,j} &= \frac{\lambda}{\alpha_{i,j}} \quad \text{and} \\
\gamma_{i,j} &= 1 - \frac{\lambda}{(\alpha_{i,j})^2}
\end{aligned}
$$

The second term in (11) is independent of $\mathbf{W}$ and thus does not effect the optimization. So the objective function to be minimized for computing the transformation matrix is given by

$$
\mathbf{J}(\mathbf{W}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i,j}(q_{i,j}(\mathbf{W}) - \beta_{i,j}d_{i,j}^{h})^2
\tag{12}
$$

Next we will describe the algorithm which can be used for minimization of functions of the above form.

*B. Iterative Majorization Algorithm*

The iterative majorization algorithm [21][22][23] is used to minimize the objective function (12) to solve for the transformation matrix $\mathbf{W}$. The central idea of the majorization method is to replace iteratively the original function $\mathbf{J}(\mathbf{W})$ by an auxiliary function $g(\mathbf{W},\mathbf{V})$. The auxiliary function, also known as the majorization function of $\mathbf{J}(\mathbf{W})$ is simpler to minimize than the original function. It can be shown that the majorization function for $\mathbf{J}(\mathbf{W})$ is given by

$$
\begin{aligned}
g(\mathbf{W},\mathbf{V}) &= \text{Tr}(\mathbf{W}^T\mathbf{A}\mathbf{W}) + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i,j}(d_{i,j}^{h})^2 \\
&- 2\text{Tr}(\mathbf{V}^T\mathbf{C}(\mathbf{V})\mathbf{W})
\end{aligned}
\tag{13}
$$

Here Tr denotes the matrix trace and $\mathbf{A}$ is given by

$$
\mathbf{A} = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i,j}(\bar{\phi}_i^l - \bar{\phi}_j^h)(\bar{\phi}_i^l - \bar{\phi}_j^h)^T
\tag{14}
$$

The term $\mathbf{C}(\mathbf{V})$ is given by

$$
\begin{aligned}
\mathbf{C}(\mathbf{V}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}(\mathbf{V})(\bar{\phi}_i^l - \bar{\phi}_j^h)(\bar{\phi}_i^l - \bar{\phi}_j^h)^T; \\
c_{i,j}(\mathbf{V}) &= \begin{cases} \lambda d_{i,j}^h / q_{i,j}(\mathbf{V}); & q_{i,j}(\mathbf{V}) > 0 \\ 0; & q_{i,j}(\mathbf{V}) = 0 \end{cases}
\end{aligned}
\tag{15}
$$

First, $\mathbf{W}$ is initialized to $\mathbf{W}^0$. The different steps of the algorithm are enumerated below:

1) Start iteration with t = 0.
2) Set $\mathbf{V} = \mathbf{W}^t$.
3) Update $\mathbf{W}^t$ to $\mathbf{W}^{t+1}$, where $\mathbf{W}^{t+1}$ is the solution that minimizes the majorization function and is given by

$$
\mathbf{W} = \mathbf{A}^{-1}\mathbf{C}(\mathbf{V})\mathbf{V}
\tag{16}
$$

where $\mathbf{A}^{-1}$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}$.

4) Check for convergence. If convergence criterion is not met, set $t = t + 1$ and go to step 2, otherwise stop the iteration and output the current $\mathbf{W}$.

The output of the iterative majorization algorithm is a transformation $\mathbf{W}$ which embeds the input LR and HR images to a new Euclidean space such that the inter-distances between them closely approximates the distances had both the images been of high resolution. Please refer to [23] for more details of the iterative majorization algorithm.

*C. Matching*

During matching, the feature vectors of both the LR probe and HR gallery images are first transformed using the learned transformation matrix. If $\mathbf{x}_p$ and $\mathbf{x}_g$ denote the feature vectors corresponding to an LR probe and an HR gallery image, the transformed feature vectors are given by

$$
\hat{\mathbf{x}}_g = \mathbf{W}^T \bar{\phi}^h(\mathbf{x}_g); \qquad \hat{\mathbf{x}}_p = \mathbf{W}^T \bar{\phi}^l(\mathbf{x}_p)
\tag{17}
$$

The distance between the probe and gallery image is computed as the Euclidean distance between their transformed feature vectors as follows

$$
d = |\hat{\mathbf{x}}_g - \hat{\mathbf{x}}_p| = |\mathbf{W}^T(\bar{\phi}^h(\mathbf{x}_g) - \bar{\phi}^l(\mathbf{x}_p))|
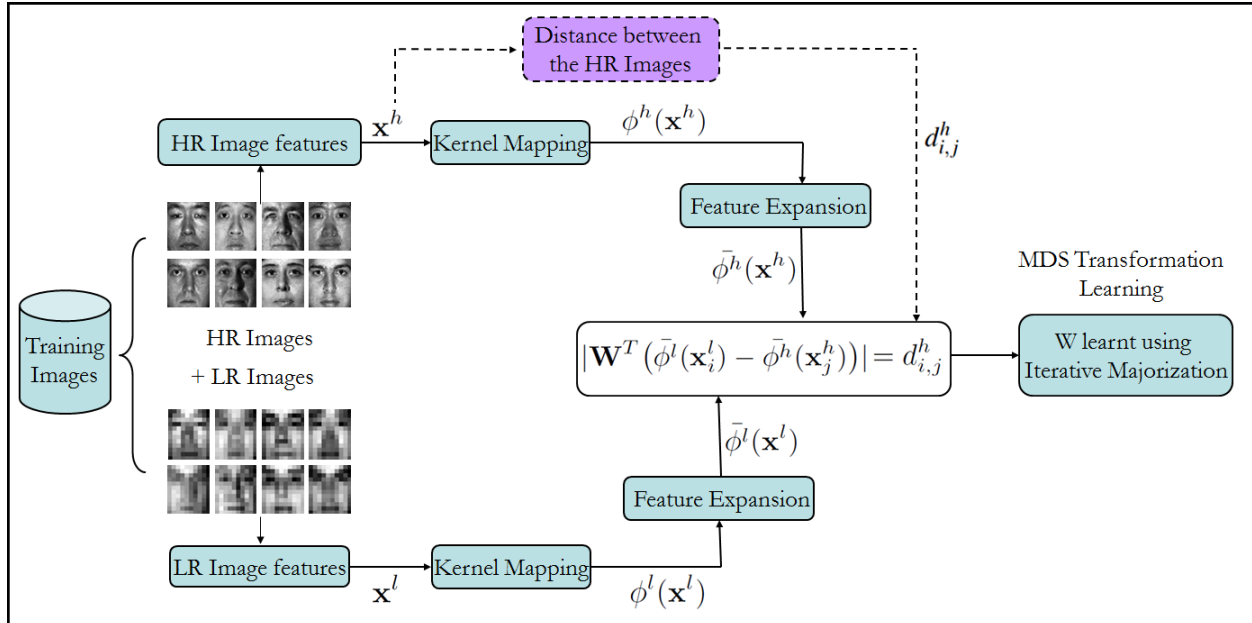\tag{18}
$$

Fig. 2. Flow chart of the proposed algorithm. The different steps are (a) feature extraction from both the HR and LR training images, (b) kernel mapping, (c) computing the expanded feature vectors as in (10). These features are used along with the distances between the HR features to compute the transformation matrix using the iterative majorization algorithm.

Since the transformation can be learned offline from training data, the algorithm is very fast and is suitable for databases of large size as can be expected from surveillance cameras. A flowchart of the proposed algorithm is shown in Fig. 2.

## V. EXPERIMENTAL EVALUATION

In this section, we describe the details of extensive experiments performed to evaluate the usefulness of the proposed approach to match LR probe images with HR gallery face images. The experiments are designed to answer the following questions

- Does the proposed approach succeed in facilitating direct comparison of LR probe images with the HR gallery images (without prior down-sampling or super-resolution step as is traditionally done)?

- How does the approach compare against the typical approach of performing super-resolution on the LR probe images to allow comparison with the HR gallery?

- How does the approach compare against standard learning-based classifiers performing

comparison in the low resolution feature space?

- How does the approach perform with different feature representations?

- How does the approach perform across different resolutions of the probe images?

- How robust is the approach to variations in the amount of training data and to different parameter choices?

*A. Data Description*

Most experiments described in this paper are performed on CMU Multi-PIE face dataset [6]. The CMU Multi-PIE face dataset is based on the popularly used CMU PIE data [24] but has a much larger number of subjects. The dataset contains images of 337 subjects who attended one to four different recording sessions which were separated by at least a month. The images are taken under different illumination conditions, pose and expressions.

For our experiments, we use the frontal images with neutral expression and varying illuminations. The images were aligned and cropped using the location of their eyes. For most experiments, images of randomly selected 100 subjects are used for training and the remaining 237 subjects for testing. Thus, there is no subject overlap across training and test sets. The aligned face images are down-sampled from the original resolution to the desired high and low resolutions for experimentation using a standard bi-cubic interpolation technique. For training, the proposed algorithm needs both HR and LR images of the same subject as side information to learn the transformation. In our experiments, given an HR training image, we down-sample and smooth it to generate the LR image of the same subject. Fig. 5 (a) and (b) shows examples of HR and LR images of three subjects.

We also perform matching experiments on the Surveillance Cameras Face Database [7]. Brief description of the dataset is provided along with the details of the experiments.

*B. Face Representation and Experiment Settings*

The proposed approach does not depend on any specific representation. In most experiments, the standard Principal Component Analysis (PCA) [25] based representation is used to perform the evaluation. FRGC training data [26] consisting of 366 face images is used to generate the PCA space. The number of PCA coefficients used to represent the face images is determined based on the number of eigenvalues required to capture $98\%$ of the total energy. We also drop coefficients

corresponding to the three largest eigenvalues for robustness to illumination variations. For the LR images, the FRGC images are first down-sampled to the resolution of the LR images and then the PCA space is generated. For all experiments in this paper, the kernel mapping $\phi$ is set to identity (i.e., $\phi(\mathbf{x}) = \mathbf{x}$) to highlight just the performance improvement due to the proposed learning approach.

Recognition is performed across illumination conditions with images from one illumination condition forming the gallery while images from another illumination condition forming the probe set. In this experiment setting, each gallery and probe set contains just one image per subject. Note that the gallery and probe sets differ in resolution with gallery being in a higher resolution than the probe images. Unless otherwise stated, all experiments report accuracy in terms of rank-1 recognition performance. The final performance is given by the average over all gallery and probe sets (there are 20 different illumination conditions in the CMU Multi-PIE dataset).

Most experiments include a baseline performance that is obtained by comparing low resolution probe images with down-sampled (originally high resolution) gallery images. This is denoted by *LR vs LR* setting in the reported results. We also report the performance obtained by directly comparing high resolution version of probe images against high resolution gallery images. This is denoted by *HR vs HR* setting in the reported results. Performance for LR vs LR setting is typically much worse than HR vs HR setting due to the loss of discriminatory information with decrease in resolution. Since these control experiments compare images of the same resolution, pairwise similarity/distance for an image pair is directly computed based on the feature representation (for PCA-based representation, it is simply the Euclidean distance between the corresponding coefficient vectors).

**Training**: For training, the weights $w_{ij}$ of the transformation matrix are initialized with random values uniformly over the range $[-1, +1]$. We now show how the objective function (error) decreases using the proposed iterative optimization process. The two terms of the objective function in (7) are separately normalized by the number of terms in each. More specifically, the first term corresponding to the distance preserving term $\mathbf{J}_{\mathrm{DP}}$ is normalized by $N \times (N - 1)$, where $N$ is the number of training images, and the second term corresponding to the class separability term $\mathbf{J}_{\mathrm{CS}}$ is normalized by the number of images of the same class in the training. The variation of the combined average of the two with iterations is shown in Fig. 3 (left). The

value of the parameter $\lambda$ is set at $0.5$. From the figure, we observe that the error decreases quite quickly and the algorithm converges in around $10$ iterations.

For all the experiments (unless otherwise stated), HR gallery images are of resolution $36 \times 30$ while the LR probe images are of resolution $12 \times 10$. Examples of HR are LR images at these resolutions are shown in Fig. 5 (a) and (b) respectively. For choosing the resolution of the HR images, we computed the recognition accuracy for different resolutions of the gallery images. Fig. 3 (right) shows that the recognition accuracy increases till about resolution $36 \times 30$ and then it saturates. So after a certain stage, increasing the resolution produces practically no improvement. Thus we have taken $36 \times 30$ as the gallery resolution for all the experiments. Though the recognition performance may not improve much with increasing the resolution beyond $36 \times 30$ using PCA coefficients, the recognition accuracy can possibly improve with higher resolutions when other representations such as LBP, etc. are used. Since most of the experiments performed in the paper use PCA coefficients as features, so this analysis is provided for PCA coefficients only. A similar experiment was performed in [2] to determine the optimal resolution of the HR gallery images. Unless otherwise stated, the weighting parameter $\lambda$ is set at $0.5$ and weights $w_{ij}$ are initialized in the same manner for all experiments in this paper.
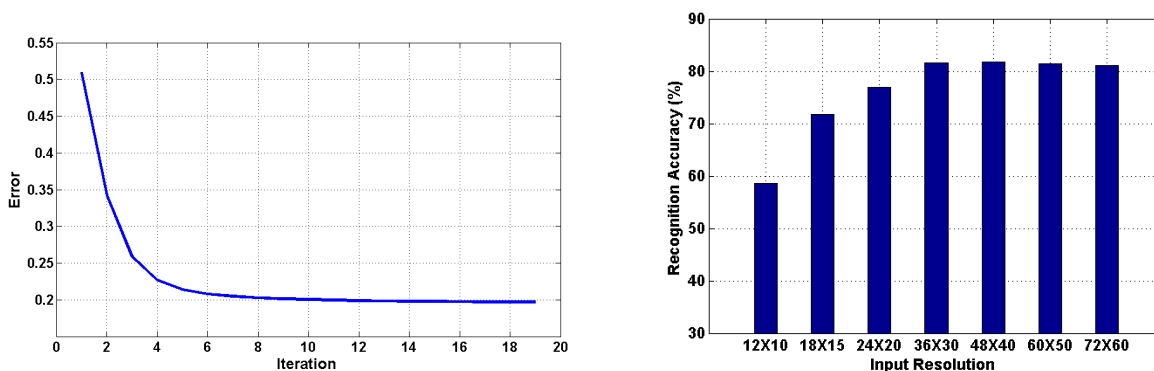


Fig. 3. (Left) Variation of normalized error with iterations during training; (Right) Recognition accuracy for different resolutions of the input images.

Fig. 4 (Left) shows the distribution of the genuine and impostor similarity scores for the LR vs LR scenario. The Multi-PIE dataset which contains images of subjects under different illumination conditions has been used to generate these distributions. The genuine scores have

been generated from images of the same subject under different illuminations, while the imposter scores have been generated from images of different subjects. Both the distributions correspond to $56880$ genuine scores and $13423680$ impostor scores. We see that the loss of discriminating information due to reduced resolution results in considerable overlap of the genuine and impostor score distributions. The corresponding distribution after the proposed MDS-based transformation is shown in Fig. 4 (Right). To measure the separability of the match and non-match distributions, we compute the $d'$ for the two cases as follows [27]

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}} \tag{19}$$

Here $\mu_1$ and $\mu_2$ denote the means and $\sigma_1$ and $\sigma_2$ denote the standard deviations of the genuine and imposter distributions. The $d'$ values before and after the proposed transformations are $1.85$ and $3.47$ respectively. Thus, we observe that the proposed approach transforms the images such that there is considerably more separation between the two distributions which results in improved recognition accuracy.
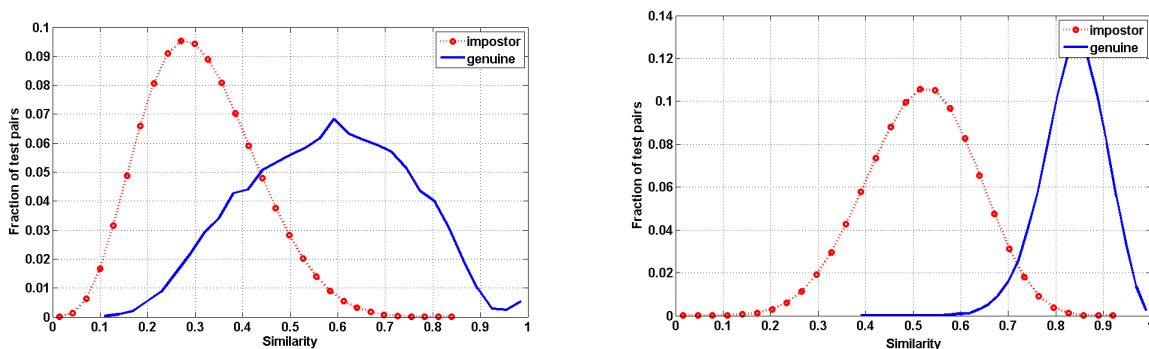


Fig. 4.  (Left) Genuine and impostor similarity score distributions for the LR vs LR scenario ($d' = 1.85$). (Right) Genuine and impostor similarity score distributions obtained using the proposed approach ($d' = 3.47$). As desired the two distributions in this case are more separated resulting in improved matching performance.

*C. Performance Comparison with Super-resolution Techniques*

For matching an LR probe with an HR gallery image, one of the most commonly used approach is to first obtain a higher resolution image from the LR probe image using a super-resolution technique which is then used for matching with the HR gallery. For comparison, we use three

different super-resolution techniques to obtain HR images from the input LR probe images. The three techniques are briefly described as follows

1) *Bicubic Interpolation* (SR1): Here the output pixel value is just a weighted average of pixels in the nearest 4-by-4 neighborhood. A standard MATLAB interpolation function is used for this technique.

2) *Sparse Representation based super-resolution* (SR2) [14]: Here, the different patches of the HR image are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal atoms. The principle of compressed sensing is used to correctly recover the sparse representation from the down-sampled input image. For our experiments, we have used the code and the pre-trained dictionary available from the author's website [28].

3) *Regression based method* (SR3) [15]: Here, the basic idea is to learn a mapping from input LR images to target HR images from example image pairs using kernel ridge regression. To remove the blurring and ringing effects around strong edges because of the regression, a natural image prior which takes into account the discontinuity property of images is used for post-processing. Code available on the author's website is used for this technique [29].

Fig. 5(c), (d) and (e) shows examples of HR images created using the three super-resolution techniques described above. The input LR images are shown in Fig. 5(b) while the actual HR images are shown in Fig. 5(a). Fig. 6 shows the recognition performance of the proposed approach along with those of the three different super-resolution approaches on images from the Multi-PIE dataset. From the recognition performance in Fig. 6, we observe that both standard LR matching and bicubic interpolation performs poorly with $58.5\%$ and $53.9\%$ average rank-1 accuracy respectively. The recognition performance of methods SR2 and SR3 are $60.58\%$ and $61.81\%$ respectively. Though the two state-of-the-art super-resolution methods improve the matching performance over LR vs LR setting, the improvement is not significant. This may be partly attributed to the fact that these super-resolution techniques are not customized for a face matching application. As desired, the proposed approach performs significantly better than all the super-resolution techniques and its performance is very close to the HR vs HR setting.

Fig. 5. (a) Comparison with super-resolution algorithms; (a) High resolution images of size $36 \times 30$; (b) Corresponding LR images of size $12 \times 10$; (c) Output images after bicubic interpolation; (d) Output images obtained using sparse-representation based super-resolution [14]; (e) Output images obtained using regression-based super-resolution [15].

## D. Performance Comparison with Classifier-based Approach

Another approach to compare LR probe images with HR gallery images is to down-sample the gallery images to the resolution of the probe images and then perform matching with the aid of a learning-based classifier. We use SVM-based classifier for this comparison. We follow the approach proposed in [30] and train SVM to classify an image pair as belonging to either intra-personal or extra-personal category. For rigorous evaluation, we use linear and RBF kernels for this experiment. SVMs are trained using the same sequestered training images of 100 subjects as done for the proposed approach. The performance of RBF kernel-based SVM depends on gamma value, which guides the smoothness of decision boundaries. A gamma value of (1/number-of-features) is typically used as default in most available SVM distributions as is the case with libsvm package used in our experiments. Along with using the default gamma value, we did a linear search for the optimal gamma value in the range $[0.01 - 10.0]$. The results of the
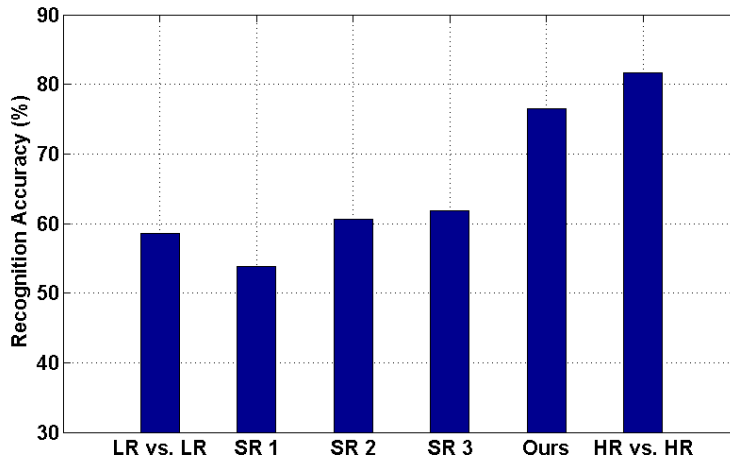
Fig. 6. Recognition performance comparison with super-resolution algorithms; (from left to right) LR vs LR setting; super-resolution using bicubic interpolation (SR1); super-resolution using sparse representation [14] (SR2); super-resolution using regression-based method [15] (SR3); proposed approach and HR vs HR setting.

experiments are shown in Fig. 7. The three SVM performances correspond to linear SVM (SVM 1), SVM with RBF kernel-default gamma (SVM 2) and SVM with RBF kernel-best performance obtained in a linear search for best gamma value (SVM 3). We observe a recognition accuracy of $61.79\%$ with default gamma value and the best recognition accuracy of $65.15\%$ (achieved at gamma = $4.0$). As can be seen from the results, though the SVM-based classification in low resolution domain does improve performance over the baseline (LR vs LR) using RBF kernel with certain gamma values, the resulting performance is still significantly lower than the performance of the proposed approach.

Fig. 7 also shows the performance obtained on the Multi-PIE data using Fisherfaces [31] which uses class information. The subspace for Fisherfaces is generated using the same sequestered training images of $100$ subjects as done for the proposed approach. The performance using Fisherfaces is $63.93\%$ as compared to $76.54\%$ using the proposed approach.

*E. Performance Analysis for Different Face Representations*

Though we have used PCA coefficients as face representation in all our experiments so far, the proposed approach is quite general and can be used with different kinds of features. Here
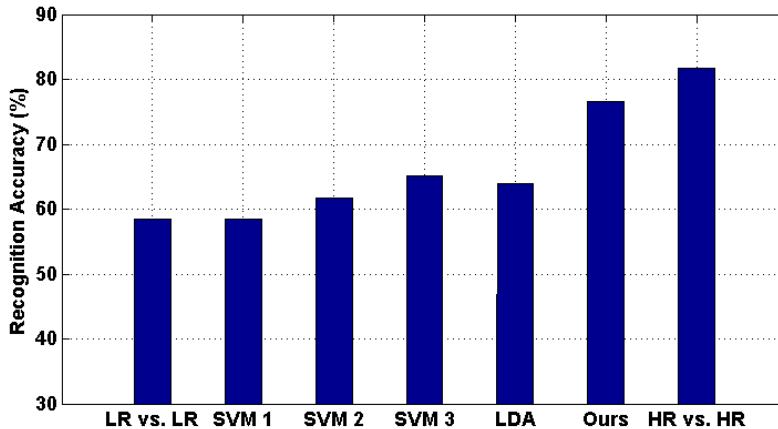
Fig. 7. Recognition performance comparison with SVM-based classification and Fisherfaces in low resolution; (from left to right) LR vs LR matching; SVM 1: SVM with linear kernel; SVM 2: SVM with RBF kernel (default gamma); SVM 3: SVM with RBF kernel (best performance obtained in a linear search for best gamma value achieved at $\gamma = 4.0$); Fisherfaces; proposed approach and HR vs HR matching.

we test the algorithm with two different kind of features which have been recently used for face recognition.

1) *Linear subspace based feature* [32] (LSBF): In this representation, the image vector is first written in terms of a shape-albedo matrix and the light source direction. An image pixel $h$ is represented as

$$h = \max(\rho \, \mathbf{n}_{3\times1}^{T} \, \mathbf{s}_{3\times1}, 0) \tag{20}$$

where $\rho$ is the albedo and $\mathbf{n}$ is the unit surface normal vector at the pixel, and $\mathbf{s}$ specifies the direction and magnitude of the light source. By stacking all the $d$ image pixels into a column vector, we have

$$\mathbf{h}_{d\times1} = [h_1, h_2, \cdots, h_d]^T = \max\left([(\rho_1 \, \mathbf{n}_1^T)\mathbf{s}, \cdots, (\rho_d \, \mathbf{n}_d^T)\mathbf{s}]^T, 0\right) = \max\left(\mathbf{T}_{d\times3}\mathbf{s}_{3\times1}, 0\right)$$

Here the matrix $\mathbf{T}$ encodes the product of albedos and surface normal vectors of all the image pixels and is termed shape-albedo matrix. Assuming that faces belong to the class of Linear Lambertian objects, the shape-albedo matrix can in turn be written as a linear combination of its basis matrices. The coefficients of the linear combination are used as

the representation for performing face recognition. Readers are directed to [32] for further details of the algorithm.

2) *Random Projection* [33] (RP) : Recently, random projection has emerged as a powerful method for dimensionality reduction that preserves the structure of the data without introducing significant distortion. Suppose the feature extraction process is given by

$$\mathbf{y} = R\mathbf{x} \tag{21}$$

where $\mathbf{x}$ is the input image, $\mathbf{y}$ is the output feature and $R$ denotes the transformation matrix. In RP, the entries of the transformation matrix $R$ are independently sampled from a zero-mean normal distribution, and each row is normalized to unit length. Readers are directed to [33] for the details of the algorithm.

TABLE I

PERFORMANCE WITH DIFFERENT FACE REPRESENTATIONS.

|  | LR vs. LR | HR vs. HR | Proposed Approach |
|---|---|---|---|
| LSBF | 68.63% | 87.23% | 80.52% |
| RP | 61.10% | 81.45% | 71.55% |

Table 1 shows the recognition performance of the approach with these features. For this experiment, 100 coefficients are chosen for both LR and HR images for the LSBF feature representation. The output feature dimension for RP is also chosen to be 100 for both LR and HR images. From the table, we see that the proposed approach performs significantly better than LR vs. LR scenario for both kinds of feature representations thus highlighting the effectiveness of the proposed MDS-based approach for different representations.

### F. Performance Analysis For Different Probe Resolutions

In this experiment, we evaluate if the good performance of the proposed approach is consistent over a wide range of resolutions of the probe images. For this experiment, HR images of resolution $48 \times 40$ pixels are used for gallery and the resolution of the probe images is varied. We perform experiments with four different resolutions of the probe images, namely $15 \times 12$, $12 \times 10$, $10 \times 8$ and $8 \times 6$. Fig. 8(a) shows examples of the HR gallery images, and Fig. 8(b), (c),

(d) and (e) show the corresponding images for the four different probe resolutions in decreasing order of resolution.



Fig. 8. Examples images of different resolution. (a) HR gallery images of resolution $48 \times 40$; Probe images of resolution (b) $15 \times 12$ pixels; (c) $12 \times 10$ pixels; (d) $10 \times 8$ pixels and (e) $8 \times 6$ pixels.

The performance of the proposed algorithm along with the performance of LR vs LR matching is shown in Fig. 9. We observe that the proposed approach can significantly improve the performance over LR vs LR image matching till probe images of resolution $12 \times 10$ pixels. The approach does improve matching performance even for resolution of $8 \times 6$ pixels but the improvement is not that significant. This may be due to the fact that LR images at such low resolution do not have enough discriminibility to help learn the desired mappings.

In this experiment, the proposed method is trained differently for different probe resolutions. In real applications where the actual resolution is unknown, a rough estimate of the resolution can be determined based on the distance between the eye centers, as is done for the Surveillance Quality Images used in Section V-H.
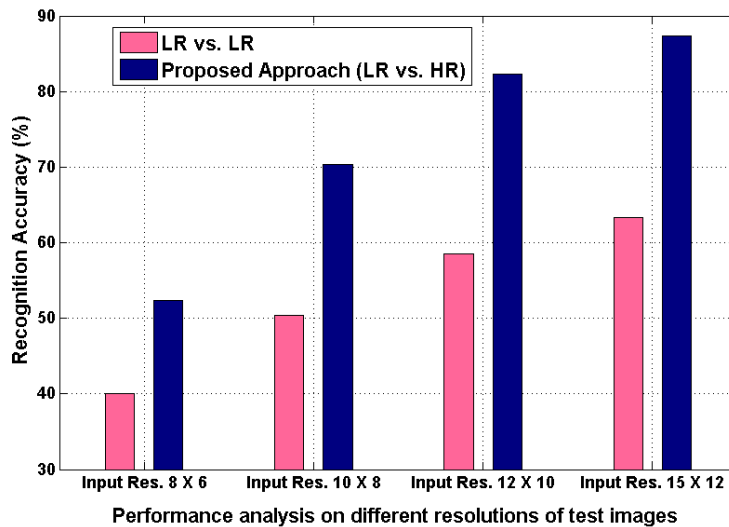
Fig. 9. Recognition performance of the proposed approach with varying probe resolutions of $8 \times 6$, $10 \times 8$, $12 \times 10$ and $15 \times 12$ respectively. Recognition performance of LR vs LR setting is also shown.

## G. Algorithm Analysis with Different Parametric Choices

We now evaluate the robustness of the proposed MDS-based algorithm with different parametric choices. Specifically, we analyze the effect of varying output dimension, weighting factor $\lambda$ and amount of training data. As before, the following analysis is performed using HR gallery images of resolution $36 \times 30$ pixels and LR probe images of resolution $12 \times 10$ pixels.

*1) Performance with varying output dimension and $\lambda$:* From Fig. 10 (right), we see that the performance of the proposed algorithm initially improves with increase in output dimension and then remains almost constant for dimension greater than around 20. For all our experiments, we have used $m = 30$. The parameter $\lambda$ determines the relative effect of the distance preserving and the class separability on the total objective function. Higher value of $\lambda$ implies a higher weight to the structure preserving term as compared to the class separability term and vice-versa. From Fig. 10 (left), we see that both the terms contribute to the good performance of the proposed algorithm. In all experiments, we have used $\lambda = 0.5$. In both the figures, the horizontal red (dashed) line represents recognition performance obtained in LR vs LR setting.

*2) Performance with different amounts of training data:* Here we analyze the performance of the proposed learning algorithm with varying amount of training data on the Multi-PIE dataset.
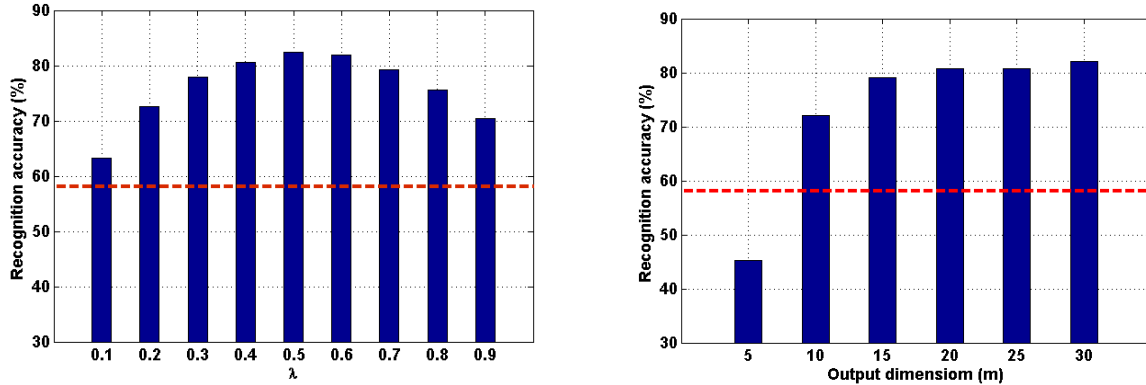
Fig. 10. Performance analysis of the proposed algorithm with varying $\lambda$ (Left) and varying output dimension (Right). The horizontal red (dotted) line represents recognition performance obtained in LR vs LR setting.

For this experiment, we vary the number of subjects used in the training set from $10$ to $100$ in steps of $10$. The number of test subjects is fixed to be $237$. The values of the other parameters used are $m = 30$ and $\lambda = 0.5$ and they are kept fixed during this experiment. The performance of the proposed approach for different number of subjects used for training is shown in Fig. 11. We see that even for number of subjects as low as $20$, the proposed approach performs better than LR vs. LR matching ($58.5\%$ average rank-1 accuracy). The recognition accuracy increases quite fast and saturates above $80$ number of subjects. In all the experiments, we have used $100$ subjects for training.

*H. Evaluation on Surveillance Quality Images*

We now test the usefulness of the proposed approach on the Surveillance Cameras Face Database [7] which contains images of $130$ subjects taken in uncontrolled indoor environment using five video surveillance cameras of various qualities. As in typical commercial surveillance systems, the database was collected with the camera placed slightly above the subject's head and also, the individuals were not required to look at a fixed point during the recordings, thus making the data even more challenging. The gallery images were captured using a high-quality photo camera. We use images from four out of the five surveillance cameras since they are close to frontal, resulting in a total of $520$ probe images. Fig. 12 shows sample gallery (top row) and probe images (bottom row) of a few subjects.
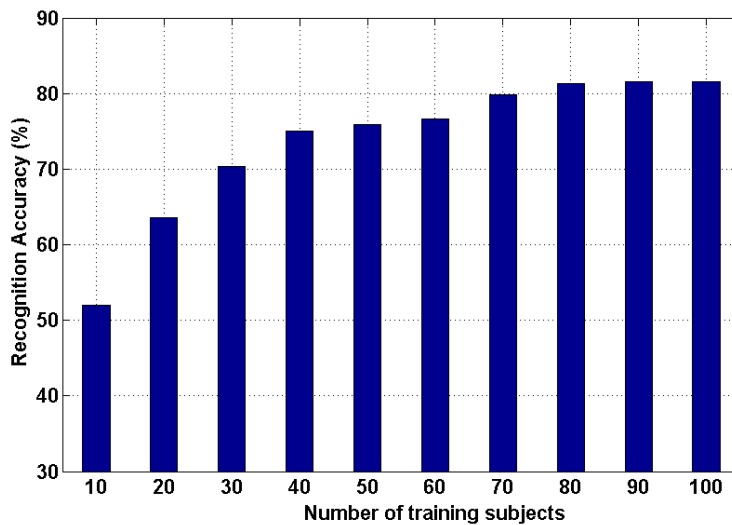
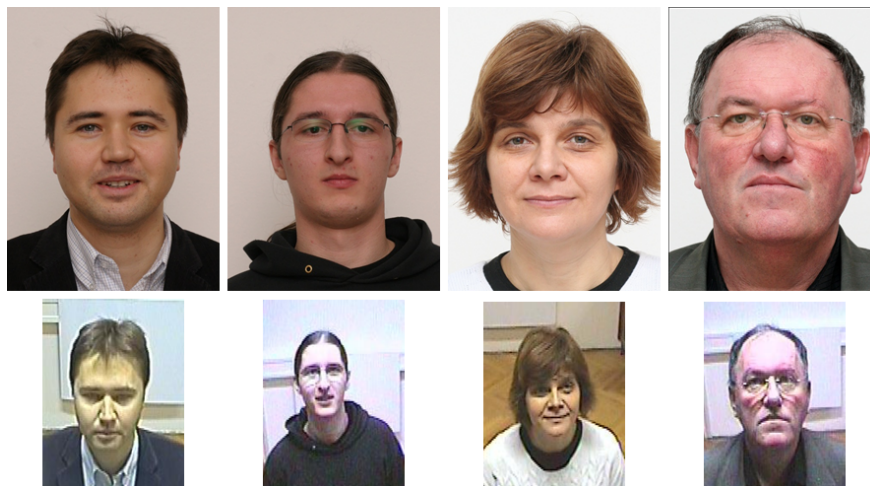Fig. 11. Performance analysis of the proposed algorithm with varying amount of training data.



Fig. 12. First Row: Example gallery images; Second Row: Example probe images of the same subjects as the gallery images [7].

For the experiment, we use images of randomly picked 50 subjects for training and the remaining 80 subjects for testing. Thus there is no identity overlap between the training and test sets. The experiment is repeated 10 times with different random sampling of the subjects. Fig. 13 shows the Cumulative Match Characteristic (CMC) for this experiment using PCA coefficients and LBP codes [34] as the input features. Error bars indicate the variation in performance for

different runs of the experiment. The number of PCA coefficients is determined based on the number of eigenvalues required to capture $98\%$ of the total energy. For the LBP feature, the face is divided into $3 \times 3$ regions and uniform LBP with $8$ neighborhood is computed. Here, we use $3 \times 3$ regions for computing the LBP features since the resolution of the probe images does not allow for bigger regions. We see that the proposed approach performs significantly better than the baseline LR vs LR setting both for PCA features as well as for more powerful representation of faces like LBP. As before, the baseline approach involves down-sampling the HR gallery images to the same resolution as the LR probe images, and then comparing the corresponding features.

Fig. 13 also shows the performance of Fisherfaces on this data. Though the Fisherfaces subspace is learned using the same subset of LR - HR images, the proposed approach performs considerably better than Fisherfaces.
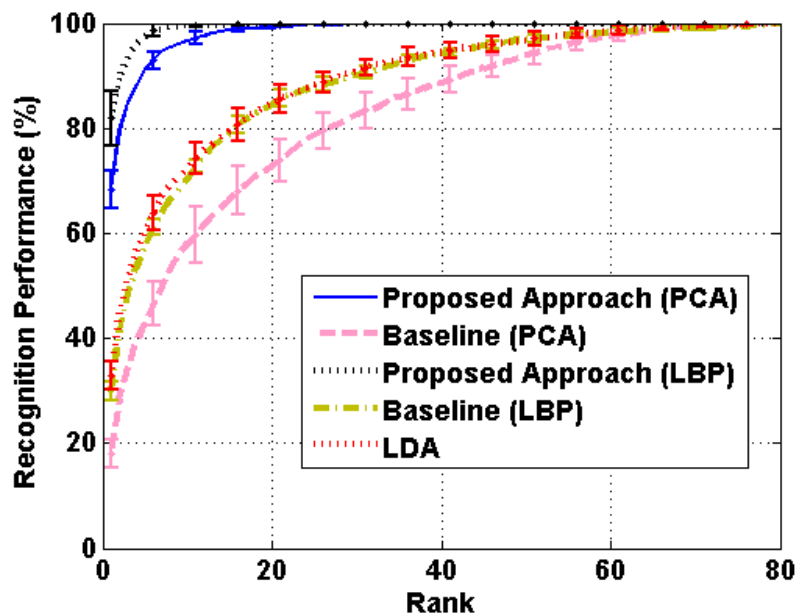


Fig. 13.   CMC obtained using the proposed approach and the baseline approach (for both PCA coefficients and LBP codes as features) on Surveillance Cameras Face Database [7]. Performance using Fisherfaces is also shown for comparison.

## VI. Summary and Discussion

In this paper, we proposed a novel approach for improving the recognition performance of low resolution images as usually obtained from surveillance cameras using Multidimensional Scaling. The main idea is to learn a transformation of the low resolution probe images and the high resolution gallery images to a common space such that the distance between them approximates the distances had these images been of good resolution. The two mappings are learned simultaneously using iterative majorization technique from HR training images. As shown in the experimental section, the learned transformations result in better separation of genuine and impostor score distributions that is the basis of the improved performance of the proposed algorithm. The desired transformations are learned during off-line training. The matching process involves simple projection of the test images using the learned transformations.

The proposed approach makes it possible to robustly match faces across resolution differences without the need of any down-sampling or super-resolution step. Extensive experimental evaluation shows the usefulness of the proposed approach. The proposed approach is compared against several approaches that include 1) simple down-sampling of gallery images for matching, 2) super-resolution of probe images using three different approaches, and 3) SVM-based classification approach. Results of the conducted experiments also show that the proposed approach can be used with different facial characterizations and is useful across a range of low resolutions. Though all the experiments have been performed on facial images, the proposed approach is very general and can be applied to any domain where such learning is feasible.

### Acknowledgement

### References

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

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

[3] K. Jia and S. Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition," in *IEEE International Conf. on Computer Vision*, 2005, pp. 1683–1690.

[4] S. Baker and T. Kanade, "Hallucinating faces," in *Fourth International Conference on Automatic Face and Gesture Recognition*, March 2000.

[5] A. Chakrabarti, A. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel pca-based prior," *IEEE Trans. on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.

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

[7] M. Grgic, K. Delac, and S. Grgic, "Scface - surveillance cameras face database," *Multimedia Tools and Applications Journal*, 2009.

[8] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution facial images," in *IEEE International Conf. On Biometrics: Theory, Applications And Systems*, 2010.

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

[10] C. Liu, H. Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[11] W. T. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[12] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004, pp. 275–282.

[13] W. Liu, D. Lin, and X. Tang, "Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 478–484.

[14] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[15] I. K. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, June 2010.

[16] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes, and R. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. on Image Processing*, vol. 12, no. 5, pp. 597–606, May 2003.

[17] S. Lee, J. Park, and S. Lee, "Low resolution face recognition based on support vector data description," *Pattern Recognition*, vol. 39, no. 9, pp. 1809–1812, 2006.

[18] J. Heo, M. Savvides, and B. Vijayakumar, "Advanced correlation filters for face recognition using low-resolution visual and thermal imagery," *LNCS*, vol. 3656, pp. 1089–1097, 2005.

[19] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 817–834, September 1999.

[20] O. Arandjelovic and R. Cipolla, "A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution," in *IEEE International Conf. on Computer Vision*, 2007.

[21] A. Webb, "Multidimensional scaling by iterative majorization using radial basis functions," *Pattern Recognition*, vol. 28, no. 5, pp. 753–759, May 1995.

[22] J. d. Leeuw, *Recent Developments in Statistics*, North-Holland, Amsterdam, 1977, ch. Applications of convex analysis to multi-dimensional scaling, pp. 133–145.

[23] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer, Second Edition, New York, NY, 2005.

[24] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[25] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neurosicence*, vol. 3, no. 1, pp. 71–86, 1991.

[26] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.

[27] J. G. Daugman and G. O. Williams, "A proposed standard for biometric decidability," in *CardTechSecureTech*, 1996, pp. 223–234.

[28] "http://www.ifp.illinois.edu/ jyang29/."

[29] "http://www.mpi-inf.mpg.de/ kkim/."

[30] P. J. Phillips, "Support vector machines applied to face recognition," in *Proceedings of Conference on Advances in Neural Information Processing Systems (NIPS)*, 1998, pp. 803–809.

[31] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997.

[32] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs, "Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 230–245, February 2007.

[33] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and D. W. Jacobs, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, February 2009.

[34] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, December 2006.



**Soma Biswas** is currently working as a Research Assistant Professor at the University of Notre Dame. Her research interests are in signal, image, and video processing, computer vision, and pattern recognition. She received the B.E. degree in electrical engineering from Jadavpur University, Kolkata, India, in 2001, the M.Tech. degree from the Indian Institute of Technology, Kanpur, in 2004, and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, in 2009.

**Kevin W. Bowyer** currently serves as Schubmehl-Prein Professor and Chair of the Department of Computer Science and Engineering at the University of Notre Dame. He is an IEEE Fellow, a Golden Core member of the IEEE Computer Society, previously served as Editor-In-Chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence, and currently serves as Editor-in-Chief of the IEEE Biometrics Compendium, the first IEEE virtual journal. Professor Bowyer is the founding General Chair of the IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS) conference series, served as a Program Chair for the 2011 IEEE International Conference on Automated Face and Gesture Recognition, and as General Chair for the 2011 International Joint Conference on Biometrics.

Professor Bowyer's recent research efforts focus on problems in biometrics and in data mining. Particular contributions in biometrics include algorithms for improved accuracy in iris biometrics, studies of basic phenomena in iris biometrics, studies involving identical twins, face recognition using three-dimensional shape, 2D and 3D ear biometrics, advances in multi-modal biometrics, and support of the governments Face Recognition Grand Challenge, Iris Challenge Evaluation, Face Recognition Vendor Test 2006 and Multiple Biometric Grand Challenge programs. His paper Face Recognition Technology: Security Versus Privacy, published in IEEE Technology and Society, was recognized with an "Award of Excellence" from the Society for Technical Communication in 2005. Professor Bowyer earned his PhD in Computer Science from Duke University.

**Patrick J. Flynn** is Professor of Computer Science and Engineering and Concurrent Professor of Electrical Engineering at the University of Notre Dame. He received the B.S. in Electrical Engineering (1985), the M.S. in Computer Science (1986), and the Ph.D. in Computer Science (1990) from Michigan State University, East Lansing. He has held faculty positions at Notre Dame (1990-1991, 2001-present), Washington State University (1991-1998), and Ohio State University (1998-2001). His research interests include computer vision, biometrics, and image processing. Dr. Flynn is a Senior Member of IEEE, a Fellow of IAPR, a current Associate Editor of IEEE Transactions on Information Forensics and Security and IEEE Transactions on Image Processing, a past Associate Editor and Associate Editor-in-Chief of IEEE Transactions on Pattern Analysis and Machine Intelligence, and a past associate editor of Pattern Recognition and Pattern Recognition Letters. He has received outstanding teaching awards from Washington State University and the University of Notre Dame.