Low Resolution Face Recognition Across Variations in Pose and Illumination

Sivaram Prasad Mudunuri and
Soma Biswas, *Senior Member, IEEE*

Abstract—We propose a completely automatic approach for recognizing low resolution face images captured in uncontrolled environment. The approach uses multidimensional scaling to learn a common transformation matrix for the entire face which simultaneously transforms the facial features of the low resolution and the high resolution training images such that the distance between them approximates the distance had both the images been captured under the same controlled imaging conditions. Stereo matching cost is used to obtain the similarity of two images in the transformed space. Though this gives very good recognition performance, the time taken for computing the stereo matching cost is significant. To overcome this limitation, we propose a reference-based approach in which each face image is represented by its stereo matching cost from a few reference images. Experimental evaluation on the real world challenging databases and comparison with the state-of-the-art super-resolution, classifier based and cross modal synthesis techniques show the effectiveness of the proposed algorithm.

Index Terms—Face recognition, stereo matching, multidimensional scaling, low resolution, super resolution

1 Introduction

THE increasing use of surveillance cameras for addressing security concerns has led to increased demand for fully automatic and robust face recognition systems. The images captured by the surveillance cameras usually have poor resolution, uncontrolled pose and illumination conditions which makes the task of recognizing these faces extremely challenging. Significant attention has been devoted to addressing one or more of the different challenges like poor illumination, non-frontal pose, expression, etc. [1], [2], [3], [4]. But addressing all these challenges together is essential in many applications like recognizing faces from surveillance cameras. Recently, a learning-based approach has been proposed for matching a low-resolution (LR) non-frontal probe image under uncontrolled illumination to frontal high-resolution (HR) gallery images [5]. The approach performs quite well in matching faces across pose, illumination and resolution, but it requires the locations of several landmark locations (like corners of eyes, nose and mouth etc.) both during training and testing, which is difficult specially for low-resolution images under non-frontal pose.

The proposed approach can be considered as an improvement over [5] as it does not require localizing facial landmarks in non-frontal face images at low resolution during testing. It is only during the training stage that we need locations of different fiducial points to learn the transformation matrix. There is another major difference from the approach described in [5]. In the proposed approach, a common transformation matrix for the entire face region that can map both low-resolution probe images and high resolution gallery images into a common space is learnt using multi-dimensional scaling method during training. SIFT descriptors computed from the facial image are used as the descriptors of the face. During testing, the images are aligned based on detected eye locations and then high resolution gallery and low resolution probe images are transformed to a common output space using the learned transformation matrix. Stereo matching cost of the transformed images is used to compute the distance between the two images across pose variations. The above approach gives very good recognition performance, but it requires significant computation time since the stereo cost has to be computed between the probe and all gallery images separately.

In this work, we also develop a reference-based face recognition system to make the proposed method computationally efficient without affecting the recognition performance significantly. A stereo matching algorithm has been proposed in [6] for matching faces across pose, but in this effort, we extend it to a learning-based stereo-matching algorithm for matching faces across all the variations together, namely pose, illumination and resolution.

Extensive experiments are performed to evaluate performance of the proposed algorithms on Multi-PIE dataset [7], Surveillance Cameras Face Database [8], Multiple Biometric Grand Challenge (MBGC) database [9] and Choke Point database [10]. The main contributions of the paper (and differences from related approaches [5], [6]) are given below:

- During training, the transformation is learned for the entire face image as opposed to selected fiducial locations as in [5].
- A completely automatic learning-based stereo matching approach for matching facial images across illumination, pose and resolution.
- A computationally efficient reference-based approach for reducing the computational cost of the approach.
- Extensive experiments are conducted on real world challenging datasets to evaluate the efficacy of the proposed approaches.

The differences will be highlighted in the respective sections in the paper.

2 Related Work

In this section, we provide pointers to the relevant papers in the literature. A 3D morphable model based approach for estimating the shape and texture information is presented in [11] for matching faces across pose and illumination variations. Ho and Chellappa [1] propose a pose invariant face recognition algorithm by using Markov Random Fields. A probabilistic model based approach which can model the appearance changes by considering the local sub regions of faces for different views is proposed in [12]. Castillo and Jacobs [2] propose a window based dense stereo matching which can address large pose variations. Chai et al. [4] propose a pose invariant face recognition method by using patch based rectification. Here virtual frontal views are generated from the given non-frontal view by estimating an approximate linear transformation between the non-frontal and frontal face. Set-Theoretic Characterization based approach is proposed in [13] to address the degradation due to blur, pose and illumination. A pose normalization algorithm to handle different poses is described in [14]. A metric learning based approach that learns a discriminative latent space by using the information of both positive and negative pairs is described in [15]. Zhu et al. [16] propose a transductive subspace learning method for matching NIR-VIS facial images as a task of heterogeneous face matching. Lu et al. [17] propose a novel neighborhood repulsed metric learning method for the task of kinship verification. A Gaussian mixture model and convex optimization based metric learning approach is presented in [18].

A detailed discussion of emerging challenges involved in recognizing low-resolution facial images is presented in [19]. Baker and Kanade [20] propose an approach for learning the resolution enhancement function for frontal facial images. Nishiyama et al. [21] propose a clustering based face recognition algorithm to recognize blurred faces. Zou et al. [22] address the problem of recognizing low-resolution facial images by learning the relationship

• *The authors are with the Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012, India.*
E-mail: {sivaram.prasad, soma.biswas}@ee.iisc.ernet.in.

Manuscript received 10 Sept. 2014; revised 28 May 2015; accepted 20 July 2015. Date of publication 17 Aug. 2015; date of current version 8 Apr. 2016.
Recommended for acceptance by M. Tistarelli.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2015.2469282

between the high-resolution and low-resolution image spaces in the training phase. A linear regression model is used to learn the relationship by imposing different constraints, and a discriminative constraint is also developed for machine-based recognition purposes. The algorithms that learn coupled mappings which can map low resolution probe images and high resolution gallery faces into a unified latent space to improve the recognition accuracy can be found in [23] and [24].

The idea of reference faces is also related to the simile classifiers proposed in [25] that can measure the similarity of the given face with some reference faces. In [25], the facial part of the probe image is classified as being similar to one of the reference faces, while in the proposed approach, the relative distance between the probe image and all the reference faces is used as the feature representation.

3 Proposed Approach

Here we describe in detail the proposed approach for matching facial images across pose, illumination and resolution. The framework consist of two stages, namely the training stage for computing the transformation matrix and the testing stage. The transformation matrix is learned in the training stage from HR frontal and LR non-frontal training images. During testing, the gallery and probe images are transformed into the common space and then stereo cost between two transformed images is computed which gives the distance between the two images.

3.1 Learning the Transformation Matrix

During training, high resolution frontal images and low resolution images under non-frontal pose are used to learn the transformation matrix. Each face in the training data is represented by a collection of local descriptors computed at fiducial locations which are extracted using STASM [26] under manual supervision to correct any isolated gross error in localization of fiducial point(s). In this work, we compute rootSIFT [27] descriptors (termed as SIFT in the remaining paper) at 15 fiducial locations in the interior of the face as the face representation.

Let the transform $\mathbf{g}$ be defined by $\mathbf{g} : R^n \rightarrow R^d$, where n is the dimension of input feature vectors and d is the dimension of the transformed space. The mapping $\mathbf{g} = (g_1, g_2, \ldots, g_d)^{\mathrm{T}}$ can be expressed as a linear combination of k basis vectors as given below

$$g_i(\mathbf{f}; \mathbf{W}) = \sum_{j=1}^{k} w_{ji} \psi_j(\mathbf{f}), \tag{1}$$

where $\psi_j(\mathbf{f})$; $j = 1, 2, \ldots, k$ is a linear or non-linear function, where $\mathbf{f}$ is the input feature vector and $\mathbf{W}$ is the transformation matrix whose elements are to be computed. Our goal is to find a transform which satisfies the following two criterion: (1) distance between the feature vectors of the HR and LR images (denoted by $\mathbf{f}_i$ and $\mathbf{f}_j$ respectively) in the transformed space should be close to the distance if both the images were captured under the same controlled imaging conditions (denoted by d_{ij}); and (2) the distance between feature vectors of the same subject in the transformed space is small as compared to that between different subjects to ensure discriminability. To achieve this, we find the transformation $\mathbf{W}$ by minimizing the following objective function

$$\mathbf{J}(\mathbf{W}) = \lambda \mathbf{J_1}(\mathbf{W}) + (1 - \lambda)\mathbf{J_2}(\mathbf{W}), \tag{2}$$

where $\mathbf{J_1}(\mathbf{W})$ is the distance preserving quantity and $\mathbf{J_2}(\mathbf{W})$ is the discriminability term. The parameter λ determines the relative importance given to the distance preserving and class separability functions. The first term is given by

$$\mathbf{J_1}(\mathbf{W}) = \sum_{i=1}^{N} \sum_{j=1}^{N} (D_{ij}(\mathbf{W}) - d_{ij})^2, \tag{3}$$

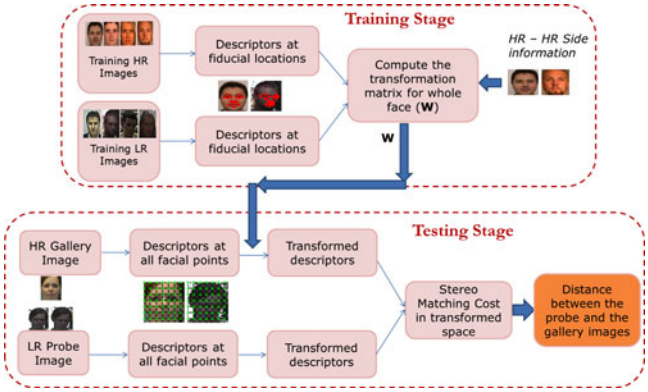

Fig. 1. Flowchart of the training and testing stages of the proposed approach.

where D_{ij} is the distance between transformed features of ith HR image and jth LR image. In this work, the second term $\mathbf{J_2}(\mathbf{W})$ which ensures class separability is given by [28]

$$\mathbf{J_2}(\mathbf{W}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(\omega_i, \omega_j) D_{ij}^2(\mathbf{W}). \tag{4}$$

Here $\delta(\omega_i, \omega_j) = 1$ if $\omega_i = \omega_j$ and 0 otherwise. The formulation above is similar to [5] with the following main differences:

- The learned transformation matrix $\mathbf{W}$ is applicable to features extracted from any part of the face as opposed to only some selected fiducial locations. So during testing, it can be used to transform the features from the entire face. This is required since we do not detect specific fiducial locations of the face during testing.
- Dimensionality reduction techniques like PCA are probably not appropriate in the proposed approach since correspondence between the fiducial locations is not assumed/maintained in the testing stage.

Finally, the transformation matrix $\mathbf{W}$ is computed by solving Eq. (2) using iterative majorization algorithm [28].

3.2 Testing

In the testing phase, the SIFT descriptors are computed at every point of the probe and gallery images (on a regular grid as shown in Fig. 1) which are then transformed to the common space using the transformation $\mathbf{W}$ learned in the training stage.

Let $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_M] \in R^{128}$ be the SIFT descriptors at M facial grid locations, then the transformed features are given by

$$\mathbf{p}^i = \mathbf{W}^T \psi(\mathbf{f}_i); \qquad i = 1, \ldots, M. \tag{5}$$

The distance between a gallery and a probe image is then computed by finding the stereo matching cost between the two images. In the proposed approach, the stereo cost is computed between transformed SIFT features of each row of HR frontal gallery and LR non-frontal probe image in the learned discriminative space. The dense four state stereo algorithm proposed by Criminisi et al. [29] is used to compute the stereo matching cost in the transformed space. We provide some details of the algorithm for completion.

The stereo algorithm includes four cumulative cost matrices namely M_{Lo}, M_{Ro}, M_{Lm} and M_{Rm}. Among these four cost matrices, M_{Lo} and M_{Ro} are designed to capture the occlusions and M_{Lm} and M_{Rm} are designed to capture the matching in left and right images respectively. The entries in all the four matrices are initialized to $+\infty$ except in the right occluded cumulative cost matrix M_{Ro} where:

$$\mathrm{M}_{\mathrm{Ro}}(i, 0) = i\alpha; \qquad i = 0, 1, 2, \ldots\ldots (q - 1). \tag{6}$$

Where q is the number of features computed from one row of the input image. The four cumulative cost matrices are computed by using dynamic programming algorithm as given in the following recursion:

$$\mathrm{M_{Lo}}(l,r) = \min \begin{cases} \mathrm{M_{Lo}}(l,r-1) + \alpha \\ \mathrm{M_{Lm}}(l,r-1) + \beta \\ \mathrm{M_{Rm}}(l,r-1) + \beta \end{cases} \tag{7}$$

$$\mathrm{M_{Lm}}(l,r) = \mathrm{M}(l,r) + min \begin{cases} \mathrm{M_{Lo}}(l,r-1) + \beta' \\ \mathrm{M_{Lm}}((l,r-1) + \gamma \\ \mathrm{M_{Rm}}(l,r-1) \\ \mathrm{M_{Ro}}(l,r-1) + \beta'. \end{cases} \tag{8}$$

Here M(l,r) is the cost of matching the transformed feature descriptors corresponding to the lth and rth grid locations in the left and right scan lines respectively. Here, l and r varies from 0 to $q-1$. M_{Rm} and M_{Ro} are symmetric. The matching cost M(l,r) is calculated as follows:

$$\mathrm{M}(l,r) = \frac{1}{2} \frac{\sum_{\delta \in \Omega} \| (\mathbf{p}^1_{l+\delta} - \bar{\mathbf{p}}^1_l) - (\mathbf{p}^2_{r+\delta} - \bar{\mathbf{p}}^2_r) \|}{\sum_{\delta \in \Omega} \| \mathbf{p}^1_{l+\delta} - \bar{\mathbf{p}}^1_l \| + \sum_{\delta \in \Omega} \| \mathbf{p}^2_{r+\delta} - \bar{\mathbf{p}}^2_r \|}, \tag{9}$$

where Ω is considered as 3×3 grid patch around each feature location (l,r). The superscript 1 and 2 denotes the two images whose stereo cost is being computed. $\mathbf{p}_k$ is transformed SIFT descriptor computed at the kth grid location. The mean of a patch is denoted with the bar. In our experiments, the different parameter values used are $\alpha = 0.5$, $\beta = 1$, $\beta' = 1$ and $\gamma = 0.25$. We have experimented with different values of these parameters and have found the algorithm to be quite robust to the parameter values.

Suppose l_1 and l_2 are two scan lines in two images, then the cost of matching these two scan lines is $\mathrm{Cost}(l_1, l_2) = M_{Ro}(q-1, q-1)$. Hence the cost of matching the probe image I_1 and gallery image I_2 is given by:

$$\mathrm{Cost}(\mathrm{I}_1, \mathrm{I}_2) = \sum_{i=1}^{N_s} \mathrm{Cost}(\mathbf{p}^1_i, \mathbf{p}^2_i), \tag{10}$$

where $\mathbf{p}^1_i$ and $\mathbf{p}^2_i$ are the transformed feature vectors corresponding to the ith scan line in the probe image and gallery image respectively and N_s is the number of scan lines. Finally, the distance between the two images is computed as follows:

$$\mathrm{Distance}(\mathrm{I}_1, \mathrm{I}_2) = \min(\mathrm{Cost}(\mathrm{I}_1, \mathrm{I}_2), \mathrm{Cost}(\mathrm{I}_2, \mathrm{I}_1)). \tag{11}$$

This is done since one does not know which image is left and which one is right in practice. A flowchart of the training and testing stages of the proposed approach is given in Fig. 1.

4 Reference Based Face Recognition

The algorithm presented above in Section 3 gives very good recognition performance as will be shown in the experimental section, but the time required for computing the distance of the probe image with the gallery images is considerably high. The main computation time is required in computing the stereo cost between two images. Given that a probe image needs to be compared against all gallery images, stereo matching cost has to be computed for each gallery image separately that makes the process both slow and non-scalable. In this section, we describe a reference-based face recognition system that aims at reducing the computation time of the algorithm without significantly affecting the recognition accuracy. In this algorithm, we select a set of reference faces and every other face (from gallery or probe) is represented by its distance relative to this set of reference faces. There is no overlap between the reference subjects with the subjects in the gallery and probe data.

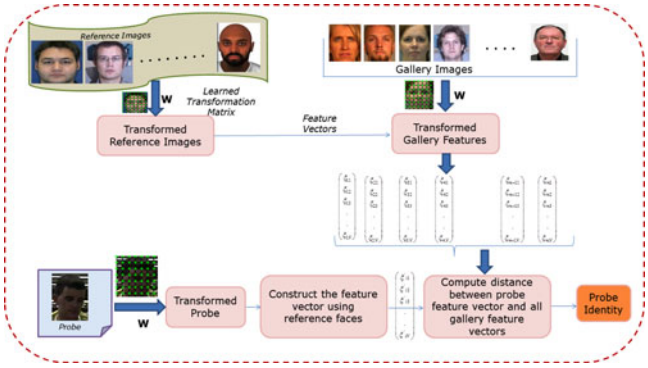

Fig. 2. Proposed reference based face recognition algorithm.

The approach using reference faces also consists of a training stage and a testing stage. In the training stage, the training HR and LR facial images are used to compute the transformation matrix **W** as in Section 3. The reference faces are HR and captured under good imaging conditions like the gallery images. In the testing stage, the features computed from the probe image, all the gallery images as well as the reference images are transformed to a common discriminative space using the matrix **W** learned during training. Note that since the gallery and reference images do not change, they can be transformed a priori to save time during testing. Each transformed gallery image is then represented by its relative distance from the transformed reference images. Let $\mathbf{f}_i$ be the feature representation of the ith gallery face and $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{ref}}$ be the feature representation of the N_{ref} reference faces. Then the gallery face is represented using the feature vector $\boldsymbol{\xi} = [\xi_{i1}, \xi_{i2}, \ldots, \xi_{iN_{ref}}]T$ where the entries in the feature vector is the stereo cost between the image and the reference images given by

$$\xi_{ij} = \mathrm{Cost}(\mathbf{W}^T \psi(\mathbf{f}_i), \mathbf{W}^T \psi(\mathbf{r}_j)); \qquad j = 1, 2, \ldots, N_{ref}. \tag{12}$$

This cost is also computed offline and so does not affect the computation time during testing. Each transformed probe image is similarly represented using the stereo cost from the set of transformed reference faces as in (12). A flowchart of the reference-based approach is shown in Fig. 2.

4.1 Computational Analysis

In the approach proposed in Section 3, the probe image had to be compared with all the gallery images, so N_g number of stereo matching computations will be required, where N_g is the number of gallery images. But in the reference-based approach, the number of stereo matching computations required for a probe image is N_{ref}, which is the number of reference images. So if $N_{ref} < N_g$, the computation time for the reference based method will be lower than the approach in Section 3.

We perform an experiment on the Multi-PIE data with 200 HR frontal gallery images and 1,000 LR non-frontal probe images (Pose 04_1) under different illumination conditions. The number of reference images is chosen as 50. It is observed from the Cumulative Match Characteristic (CMC) curve in Fig. 3 that though the rank-1 performance of the reference-based approach is not very good, the performance improves quite rapidly for higher ranks. The top ranked gallery image is the correct identity of the probe only 55 percent times but the correct identity is in the top 10 ranks for over 90 percent of probe images.

Based on these observations, we propose a modification of the reference-based approach. Using the proposed reference-based approach, we obtain distances between the probe and the gallery images. Based on these distances, the top K gallery images are picked and direct stereo matching (as described in Section 3) is performed between them and the probe image to obtain a better

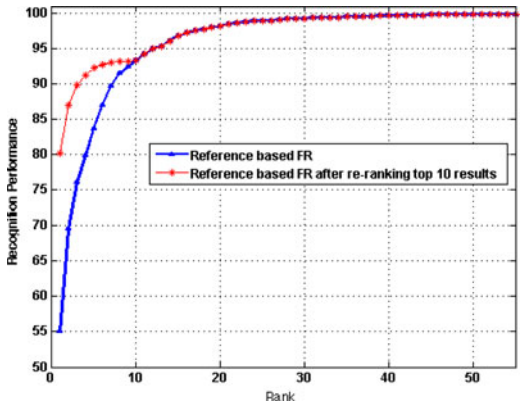

Fig. 3. Cumulative match characteristic curves for the reference-based approach and the modified reference-based approach that re-ranks the top ranked gallery images by computing the stereo cost against the probe image.

estimate of their distance leading to better accuracies at top ranks as shown in Fig. 3. K is set to 10 for this experiment. With this modification, the rank-1 accuracy improves from 55 to 80 percent at the expense of 10 extra stereo cost computations. As expected, the two CMC curves exactly overlap at later ranks. In this modified approach, the total number of stereo matching computations required is $N_{ref} + K$, and if this is less than N_g, then the computation required will be lower than the approach in Section 3. Fig. 4 illustrates the improvement in performance using the modified reference-based approach. For both query images (left), the top row shows the top 10 matches returned by the reference based approach and the bottom row shows the re-ranked results after computing the stereo matching cost with each of these 10 matches. The image with green bounding box shows the correct match.

5 Experimental Results

Extensive experiments are conducted on three datasets namely, Multi-PIE dataset [7], Surveillance Cameras Face Database [8], Multiple Biometric Grand Challenge database [9] and Choke Point database [10] to demonstrate the applicability of the proposed approach. We have used Active Shape Model-based C++ software library called STASM [26] which is freely available to detect feature locations automatically. The detections were manually verified and the incorrect locations were corrected. Note that the detected landmarks are required during training only. During testing, only the locations of the eyes are required to align the facial images.

5.1 Experiments on Multi-PIE Dataset

The Multi-PIE dataset [7] contains face images of 337 subjects that are captured under different illumination conditions and view points in four recording sessions. The probe images with four different poses namely pose 04_1, 05_0, 14_0 and 13_0 as labeled in Multi-PIE dataset are used in our experiments. The high-resolution images in frontal pose are used as gallery images. The resolution of the gallery images used in our experiments is 36×30 while the resolution of the probe images used is 18×15 (scale factor of 2).

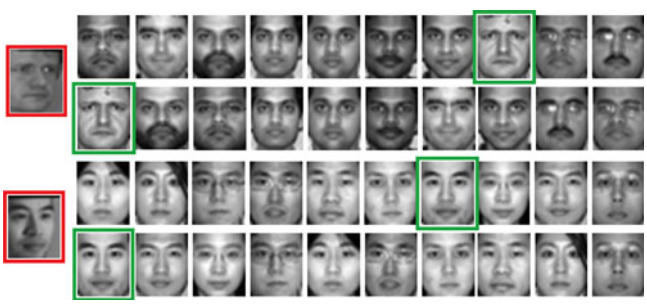

Fig. 4. Modified reference-based approach. For each query (left), top row shows the top 10 matches returned by the reference-based method, bottom row shows re-ranked result using the proposed modification.

TABLE 1
Rank-1 Recognition Performance for Four Different Probe Poses, Averaged over the Different Gallery Illuminations

Method	Pose 13_0	Pose 14_0	Pose 05_0	Pose 04_1
HR-LR (baseline) [5]	49.15%	69.87%	59.04%	35.57%
HR-LR (MDS) [5]	63.01%	76.35%	72.83%	56.32%
Stereo Baseline [6]	76.53%	85.57%	84.53%	75.36%
Semi-Coupled [30]	64.02%	72.37%	68.10%	62.93%
Dictionary Learning [31]	63.41%	72.08%	66.91%	62.37%
LMNN (min) [32]	36.74%	52.18%	40.91%	30.35%
LMNN (mean) [32]	58.10%	72.95%	62.11%	50.41%
LMNN (max) [32]	67.73%	80.64%	72.19%	59.71%
LSML ($M_{y=1}$) [33]	82.98%	90.72%	88.69%	71.36%
LSML [33]	**90.51%**	93.68%	89.17%	83.86%
SFRD + PMML [34]	77.17%	88.34%	92.72%	75.97%
GMA - LPP [35]	68.42%	78.77%	80.32%	70.13%
GMA - MFA [35]	72.40%	82.29%	84.77%	73.68%
CrossPose [36]	54.21%	73.57%	70%	60%
Proposed approach in Section 3	**88.44%**	**95.69%**	**94.81%**	**86.43%**
Proposed Reference	**80.53%**	**88.77%**	**86.27%**	**80.16%**

Recognition experiments are conducted across illumination conditions with images from one illumination forming the gallery while images from a different illumination forming the probe set. In our experiments, we use five different illumination conditions. The experiment is repeated for different pairings of illumination conditions. Reported recognition accuracy is the average rank-1 recognition performance averaged over the different illumination conditions.

Fifty randomly chosen subjects are used for training and the remaining subjects are used for testing. There is no overlap between the training and test subjects. The value of λ is set to 0.9 and d, the output dimension of MDS is set to 100 for all the experiments. The kernel mapping $\psi(x)$ is fixed to x to emphasize the performance of proposed approach, but other more appropriate mapping can also be used.

5.1.1 Performance: Different Pose, Illumination, Resolution

We start with reporting the recognition performance of the proposed approach for matching images which differ in pose, illumination and resolution.

Comparison of the proposed approach with the MDS based approach in [5] and stereo matching algorithm in [6] are also reported in Table 1. The source codes in the corresponding author's website [37], [38] are used for generating the results. In Table 1, HR - LR (baseline) indicates using SIFT descriptors but without learning the transformation in [5]. For the stereo baseline approach [6], first bilinear interpolation is applied to the LR probe images to make the resolution same as the HR gallery images before stereo matching cost is computed. Had the probe images been captured under the same pose and resolution as the gallery images , the performance (HR-HR) would have been 93.55 percent. The proposed approach not only performs considerably better than the baseline methods [5], [6], the performance is very close to the HR-HR performance. For the reference based approach, the number of reference images is taken as 50. The performance of the reference-based approach is slightly worse than the approach proposed in Section 3, it is still considerably better than the baseline approaches. Though the LSML algorithm performs better than our reference approach, LSML requires the locations of several landmark points (like corners of eyes, nose and mouth etc.) both

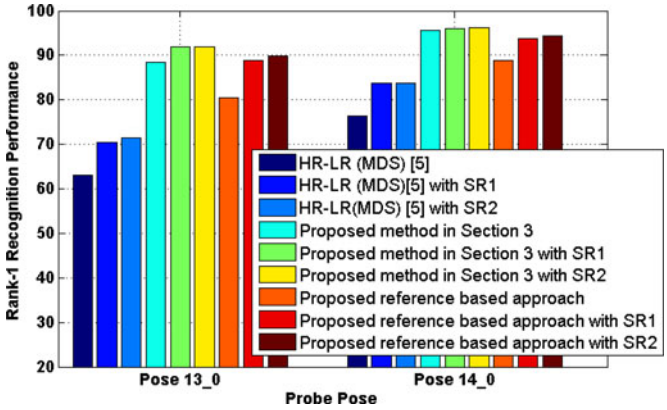

Fig. 5. Rank-1 recognition accuracy of the proposed algorithms with the two super-resolution techniques SR1 [40] and SR2 [41]. Comparison with other approaches are also shown.

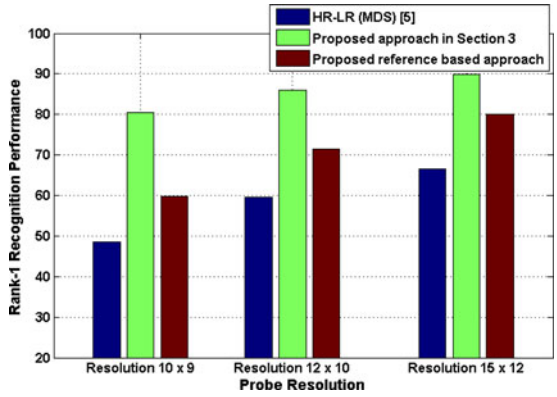

Fig. 6. Rank-1 recognition performance for different probe resolutions.

during training and testing, which is difficult especially for low-resolution images under non-frontal pose.

5.1.2 Comparison: Metric Learning Approaches

We also evaluate three state-of-the-art metric learning/classifier based algorithms: (1). Distance metric learning approaches such as Large Margin Nearest Neighbor (LMNN) [32] have been successfully used for recognizing facial images [39]. We experimented with different settings of validation, number of nearest neighbor and maximum number of iterations parameters and the minimum, maximum and mean value of rank-1 recognition performance are reported (Table 1). (2). Large Scale Metric Learning (LSML) [33] method that learns a metric from equivalence constraints based on the statistical inference perspective. The case $M_{y=1}$ represents that the metric is learned from Mahalanobis distance of the similar pairs. (3). Pairwise Multiple Metric Learning (PMML) method [34] that integrates the face region descriptors from different regions of facial images. The codes available in the author's web pages are used to generate the results for the three metric learning approaches reported in Table. 1. The proposed approach outperforms all the three approaches with the exception of LSML for one pose in which the proposed approach performs slightly worse.

5.1.3 Comparison: Cross-Modal and Dictionary Learning

Recently, cross domain image synthesis and recognition methods have demonstrated promising performance in many applications. The cross domain algorithms address the recognition/classification task across different domains. We evaluate the methods presented in [30], [31], [35] and [36]. The algorithms [30] and [31] jointly solve the coupled dictionary and common feature space learning to synthesize the cross-domain images and perform the recognition task. The Generalized Multiview Analysis (GMA) [35] approach arrives at a single linear subspace by solving a joint, relaxed quadratic constrained quadratic program (QCQP) over different feature spaces. The algorithm in [36] handles the problem of recognizing cross pose facial images by modeling a regressor with a coupled bias variance tradeoff. We use the codes available in respective author's webpage and evaluate each algorithm with different parameter settings and the best results are reported in Table 1. The proposed approach outperforms all these approaches for all poses.

5.1.4 Comparison: Super-Resolution Approaches

One of the most commonly used technique for matching a LR probe image with HR gallery image is to use super resolution on the LR probe images before matching is performed. We compare the proposed approach with two state-of-the-art SR techniques SR1 [40] and SR2 [41]. In SR1 [40], kernel ridge regression method is employed to learn a map from LR images to the desired HR images. In SR2 [41], two different dictionaries are maintained to train the LR and HR images, and the sparse representation of the LR images is used to get the corresponding HR image. The codes from the respective author's websites are used to generate the results in our experiment.

From Fig. 5, it can be observed that the proposed approaches significantly outperform the compared approaches for both probe poses and SR techniques.

5.1.5 Recognition Across Different Probe Resolutions

In this experiment, we analyze the performance of the proposed algorithm for wide range of resolutions of probe images. The gallery images are maintained at a resolution of 36×30 and the resolution of probe images are varied. We use three different probe resolutions, 15×12, 12×10 and 10×9. LR probe images of pose 05_0 are used for this experiment.

The rank-1 recognition performance of the proposed algorithms and MDS-based algorithm in [5] are presented in Fig. 6. We see that the proposed algorithms perform significantly better than the compared approach for all the probe resolutions.

5.1.6 Advantage of the Reference-Based Approach

We study the recognition performance and computational advantage of the proposed reference based approach by varying the number of reference images.

Table 2 shows the rank-1 recognition performance and runtimes of the proposed reference-based approach for different number of reference images. This experiment is conducted on the Multi-PIE data with the same experimental setting as the previous experiments for probe pose 04_1. For a probe image, after the similar gallery images are returned by the reference-based approach, the stereo matching cost is computed between the probe image and the top 10 similar gallery images which helps to significantly improve the rank-1 recognition accuracy.

TABLE 2
Rank-1 Recognition of the Modified Reference-Based Algorithm with Varying Number of Reference Images for Pose 04_1

No. of ref. images	Rank-1 Accuracy	Time (seconds)
10	74.75%	14
20	78.85%	16
30	79.81%	19
40	79.92%	21
50	80.16%	24
Proposed Section 3	86.43%	101

TABLE 3
Rank-1 Recognition of the Proposed Approach and Comparison with Existing Algorithms on SCFace and MBGC Databases

Method	SCFace [8]	MBGC [9]
MDS HR-LR [5]	61.14%	39.48%
LSML [33]	59.25%	49.15%
GMA-MFA [35]	27.0%	19.21%
Proposed Approach	**69.45%**	**50.57%**

It is observed that even with as few as 20 reference images, the proposed reference-based approach gives noticeable better performance compared to many of the existing algorithms that are reported in Table 1. Obviously there is a trade off between the number of reference images used and the recognition accuracy obtained. But the computational time decreases significantly from 101 seconds to just 16 seconds with this modified reference-based approach.

5.2 Experiments on Surveillance Camera Dataset

We now evaluate the proposed approach on real surveillance quality data obtained from the Surveillance Cameras Face Database [8]. The dataset contains images of 130 subjects captured in uncontrolled environment using five different video surveillance cameras, while the gallery images were taken using high-quality camera. We use the same experimental setting as used in [5], in which we use all the images from all the five surveillance cameras (thus there are 650 images).

As in [5], we randomly pick 50 subjects for training and use the remaining 80 subjects for testing (thus there are a total of 400 probe images) with no overlap between the train and test subjects. The experiment is repeated 10 times with different random sampling of the subjects. The Rank-1 accuracy of the proposed approach and comparisons with several other approaches for this experiment are shown in the second column of Table 3. We see that even for real surveillance quality data, the proposed approach performs significantly better than the other approaches.

5.3 Experiments on MBGC Database

The proposed algorithm is also evaluated on Multiple Biometric Grand Challenge [9] database to demonstrate its efficacy in a real world surveillance scenario. The database includes frontal images of 147 subjects and videos of the same subjects where each user is walking or performing some activity. The frontal images are taken as gallery images and the faces extracted from videos of corresponding persons are considered as probe images in our experiment. Faces present in these videos are very different from the gallery images in terms of variations due to resolution, illumination and pose. A few sample images are shown in the Fig. 7.

Gallery consists of single image per subject and probe set consists of five images per subject in our experiment. A total of 70 randomly selected subjects are used for training and the remaining subjects are used for testing, thus there is no overlap between the training and test subjects. The experiment is repeated 10 times with random selection of training and test subjects and the average Rank-1 accuracy is reported. Experimental results of proposed algorithm and comparison with state-of-the-art approaches are reported in the third column of Table 3 to demonstrate the effectiveness of our algorithm.

Fig. 7. MBGC [8] data- Top row: Sample gallery images. Bottom Row: Sample probe images of the corresponding subjects.

5.4 Experiments on ChokePoint Database

The proposed algorithm is also evaluated on the Choke Point database [10] to further demonstrate its efficacy in real world surveillance scenarios. We have followed the same protocol as that of Bhatt et al. [42] for this experiment. As in [42], images of Multi-PIE dataset are used for training and all the 29 subjects in the Choke Point database for testing. We repeated our experiment five times by randomly selecting the probe images each time and the mean accuracy is reported in Table 4. The rank-1 accuracies of all the other approaches that are reported in Table 4 are directly taken from [42]. A few sample gallery and probe images of the dataset are shown in Fig. 8. We see that for this dataset, the proposed approach performs significantly better than all the other approaches.

6 Discussion and Conclusion

We have presented a novel face recognition algorithm for matching faces across different pose, illumination and resolution. A transformation matrix is learned for the entire image in the training stage using multidimensional scaling. The cost of stereo matching between the gallery and the probe image in the transformed space is taken as the distance between the two images for computing the recognition performance. We also proposed a reference based face recognition algorithm for reducing the computational requirement. The computational time of the proposed approach is higher as compared to the other approaches. As part of our future work, we would like to explore recent fast and efficient stereo matching algorithms [43] that can potentially decrease the time required for the proposed approach. Having said that, the main advantage of the proposed approach is that there is no need to mark any fiducial locations during testing. All the other approaches in Table 1 require fiducial locations to be marked on the low-resolution

TABLE 4
Rank-1 Recognition (%) of the Proposed Approach and Comparison with Existing Algorithms on ChokePoint Database [10]

Resolution		Algorithm														
Gallery	Probe	LPQ	SIFT	E1	E2	Fusion	MDS	CTL	HR/ LR TL (LPQ)	HR/ LR TL (SIFT)	HR/LR TL LPQ+ (SIFT)	HR/ LR CT	COTS	CTL+ MDS	CTL+ COTS	Proposed
48 × 48	32 × 32	35.4	32.6	41.2	37.6	44.7	45.4	48.2	46.6	43.4	47.1	40.4	18.5	47.8	50.9	**62.7**
	24 × 24	23.2	20.4	27.4	24.8	29.5	30.2	33.1	31.6	29.1	32.8	27.1	11.8	32.6	37.2	**60.6**
	16 × 16	17.6	14.5	21.8	19.6	24.1	26.3	28.3	25.8	23.6	26.5	22.7	4.7	27.5	31.6	**54.3**
32 × 32	24 × 24	20.4	14.8	23.4	18.7	24.3	28.6	31.6	26.2	25.6	29.4	21.3	16.4	30.8	35.4	**58.4**
	16 × 16	14.6	9.6	17.3	13.4	19.6	21.9	23.1	21.1	19.2	21.8	15.6	3.5	22.5	26.0	**56.8**

Fig. 8. Example facial images of Choke Point database [10]. Top row: frontal gallery images, second row: corresponding probe images.

images during testing, which is a challenging task. The usefulness of our algorithms is justified with experiments conducted on the Multi-PIE dataset, SC Face database, MBGC database and Choke-Point database in which very good recognition performance is obtained under very low resolution of probe images and wide range of pose and illumination conditions.

REFERENCES

[1] H. T. Ho and R. Chellappa, "Pose-invariant face recognition using Markov random fields," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1573–1584, Apr. 2013.

[2] C. D. Castillo and D. W. Jacobs, "Wide-baseline stereo for face recognition with large pose variation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 537–544.

[3] J. M. Chang, M. Kirby, H. Kley, C. Peterson, B. Draper, and J. R. Beveridge, "Recognition of digital images of the human face at ultra low resolution via illumination spaces," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 733–743.

[4] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[5] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer, "Pose-robust recognition of low-resolution face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 3037–3049, Dec. 2013.

[6] C. D. Castillo and D. W. Jacobs, "Using stereo matching with general epipolar geometry for 2d face recognition across pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2298–2304, Dec. 2009.

[7] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Guide to the CMU multi-pie database," Carnegie Mellon Univ., Pittsburgh, PA, USA, 2007.

[8] M. Grgic, K. Delac, and S. Grgic, "SCface—surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, 2011.

[9] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. OToole, D. S. Bolme, K. W. Bowyer, A. Draper, Bruce, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer, "Overview of the multiple biometrics grand challenge," in *Proc. Int. Conf. Biometrics*, 2009, pp. 705–714.

[10] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Proc. Comput. Vis. Pattern Recog. Workshops*, 2011, pp. 74–81.

[11] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[12] T. Kanade and A. Yamada, "Multi-subregion based probabilistic approach toward pose-invariant face recognition," in *Proc. Int. Symp. Comput. Intell. Robot. Autom.*, vol. 22, 2003, pp. 954–959.

[13] V. Priyanka, K. Mitra, and R. Chellappa, "Blur and illumination robust face recognition via set-theoretic characterization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1362–1372, Apr. 2013.

[14] L. Ding, X. Ding, and C. Fang, "Continuous pose normalization for pose-robust face recognition," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 721–724, Nov. 2012.

[15] A. Mignon and F. Jurie, "CMML: A new metric learning approach for cross modal matching," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 1–14.

[16] J. Y. Zhu, W. S. Zheng, J. H. Lai, and S. Z. Li, "Matching NIR face to VIS face using transduction," *IEEE Trans Inform. Forensics Security*, vol. 9, no. 3, pp. 501–514, Mar. 2014.

[17] J. Lu, X. Zhou, Y. P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 331–345, Feb. 2014.

[18] J. Bohne, Y. Ying, S. Gentric, and M. Pontil, "Large margin local metric learning," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 679–694.

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

[20] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[21] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi, "Facial deblur inference to improve recognition of blurred faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 1115–1122.

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

[23] J. Shi, "From local geometry to global structure: Learning latent subspace for low-resolution face image recognition," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 554–558, May 2015.

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

[25] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 365–372.

[26] S. Milborrow and F. Nicolls. (2008). Locating facial features with an extended active shape model. *Proc. Eur. Conf. Comput. Vis.*[Online]. Available http://www.milbo.users.sonic.net/stasm

[27] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

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

[29] A. Criminisi, A. Blake, C. Rother, J. Shotton, and P. H. S. Torr, "Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 89–110, 2007.

[30] S. Wang, D. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2012, pp. 2216–2223.

[31] D. A. Huang and Y. C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2496–2503.

[32] K. Q. Weinberger and L. K. Saul, "Fast solvers and efficient implementations for distance metric learning," in *Proc. Int. Conf. Mach. Learning*, 2008, pp. 1160–1167.

[33] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2288–2295.

[34] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, "Fusing robust face region descriptors via multiple metric learning for face recognition in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 3554–3561.

[35] A. Sharma, A. Kumar, H. Daume, and D. H. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2160–2167.

[36] A. Li, S. Shan, and W. Gao, "Coupled bias-variance tradeoff for cross-pose face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 305–315, Jan. 2012.

[37] [Online]. Available: http://www.ee.iisc.ernet.in/new/people/faculty/soma.biswas/research.html

[38] [Online]. Available: http://www.cs.umd.edu/ carlos/

[39] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in Proc. *IEEE Int. Conf. Comput. Vis.*, 2009, pp. 498–505.

[40] I. K. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

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

[42] H. Bhatt, R. Singh, M. Vatsa, and N. Ratha, "Improving cross-resolution face matching using ensemble based co-transfer learning," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5654–5669, Dec. 2014.

[43] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Proc. 10th Asian Conf. Comput. Vis.*, 2010, pp. 25–38 .