# ALIGNED DISCRIMINATIVE POSE ROBUST DESCRIPTORS FOR FACE AND OBJECT RECOGNITION

*Soubhik Sanyal, Devraj Mandal, Soma Biswas*

Department of Electrical Engineering, Indian Institute of Science

## ABSTRACT

Face and object recognition in uncontrolled scenarios due to pose and illumination variations, low resolution, etc. is a challenging research area. Here we propose a novel descriptor, Aligned Discriminative Pose Robust (*ADPR*) descriptor, for matching faces and objects across pose which is also robust to resolution and illumination variations. We generate virtual intermediate pose subspaces from training examples at a few poses and compute the alignment matrices of those subspaces with the frontal subspace. These matrices are then used to align the generated subspaces with the frontal one. An image is represented by a feature set obtained by projecting its low-level feature on these aligned subspaces and applying a discriminative transform. Finally, concatenating all the features we generate the *ADPR* descriptor. We perform experiments on face and object databases across pose, pose and resolution, and compare with state-of-the-art methods including deep learning approaches to show the effectiveness of our descriptor.

***Index Terms*—** Face recognition, object recognition, pose, subspace interpolation, subspace alignment.

## 1. INTRODUCTION

Face and object recognition in uncontrolled scenarios is a challenging problem in computer vision research. Here, uncontrolled scenarios refer to variations in pose, resolution and illumination in the images. For example, images obtained from surveillance systems, usually have significant pose variations and poor resolution, thus making it difficult to match with the high resolution and frontal gallery during enrolment. Similarly, different views of objects look very different, thus making object recognition across pose also a very challenging problem. The goal of this work is to match frontal gallery images (of faces and objects) with uncontrolled probe images, which consists of images in unconstrained pose, illumination and possibly low-resolution (for surveillance scenarios). A significant amount of research has been done in the area of matching across pose [1][2][3][4]. In contrast, matching poor resolution images is a relatively lesser studied problem [5][6]. Majority of these works handle pose, resolution and illumination separately. Recent advances in deep learning in face and object matching have taken the state-of-the-art performance to a very high level [7] [8]. Wang *et al.* [9] show that even these approaches are not very robust in uncontrolled scenarios.

In this work, we propose a novel descriptor for matching faces and objects across pose, illumination and resolution, termed as Aligned Discriminative Pose Robust (*ADPR*) descriptor. Given training samples from a few pose regions, first virtual intermediate subspaces are generated. Since recognition performance in frontal pose is better than that at any other pose (even when both gallery and probe are in same pose as we show later), we propose to transform every pose to the frontal pose to perform matching, For this, frontal alignment matrices are computed for all the virtually generated subspaces, which can align the generated subspaces with the frontal subspace. The feature vector of the input image is projected onto all the aligned subspaces. Finally, a discriminative transform is learned using the training class labels. The final aligned and discriminative features are concatenated to form the *ADPR* descriptor.

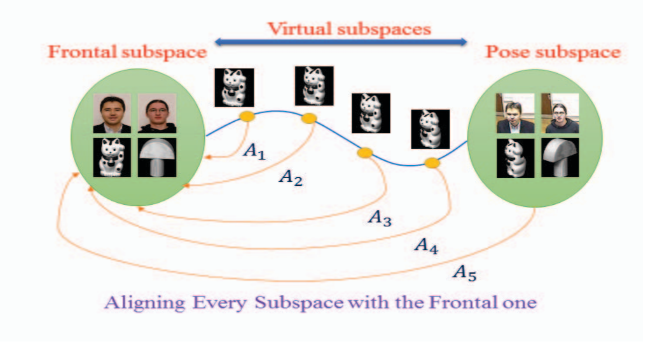


**Fig. 1**. Illustration of virtual subspace generation and frontal alignment used in the proposed descriptor computation. Images are taken from SCface [10] and COIL-20 [11] database.

The proposed work is motivated from [12] where a discriminative descriptor is proposed. The novelty/contributions of the proposed work are as follows: (1) A novel aligned discriminative pose robust descriptor is proposed which does not require separate training for different test poses or viewpoints. (2) The proposed descriptor is computationally much more efficient as compared to [12] and thus the time required to identify a probe image represented by ADPR is significantly less. (3) Extensive evaluation on two face datasets, namely Multi-PIE [13] and Surveillance Cameras Face Database [10] and two object databases, namely COIL 20 database [11] and RGB-D Object Database [14] (which also includes depth data), and comparisons with the state-of-the-art shows the effectiveness of the proposed descriptor. (4) The approach can also improve the recognition performance of the state-of-the-art deep learning architectures.

## 2. PROPOSED APPROACH

In this section, we describe in detail the computation of the proposed *ADPR* descriptor. The training stage has three parts, namely (1) generation of intermediate subspaces, (2) computation of aligning matrices and (3) computation of discriminative features. We describe each of these steps in detail in the following subsections.

### 2.1. Generation of intermediate subspaces

The goal of this work is to generate a descriptor for an image captured under any unknown pose, so that it can be used for matching across poses. For this, we consider that during training, images from some regions of the pose space, say $F_1$ to $F_K$, ($K$ is as small as two/three) are available. Suppose $p$ is the low-level feature descriptor of an image. Images captured under different poses will have different low-level descriptors. So, if we want to match images in different pose, it is better if descriptors can be computed for every pose and then the feature set is matched. Thus, we propose to represent the input image using a collection of features $\{p_1, p_2, \ldots\}$, where, $\{p_1, p_2, \ldots\}$ are the feature vectors computed if we have that image in different poses. Therefore the chances of matching two images of the same object which only differ by pose, is higher if we compare the feature sets of the two images, rather than only comparing $p$.

To generate the features at different poses, we compute virtual poses by learning the path between $F_k$ and $F_{k+1}$, by exploiting the idea of sampling on the Grassmann manifold [15]. Suppose $n_{trn}$ be the number of training images in pose $F_k$ as well as in pose $F_{k+1}$. We compute the data matrix of dimension $D \times n_{trn}$ for pose $F_k$ and $F_{k+1}$ using the low level features of the images. Here $D$ is the dimension of the low-level image feature. Next, we compute the generative subspaces $\bar{F}_k$ and $\bar{F}_{k+1} \in \mathbb{R}^{(D\times m)}$ by applying principal component analysis (PCA) on the data matrix. The space of $m$-dimensional subspaces in $\mathbb{R}^D$ can be identified with the Grassmann manifold $\mathbb{G}_{m,D}$ and thus, $\bar{F}_k$ and $\bar{F}_{k+1}$ are points on $\mathbb{G}_{m,D}$. Let $\bar{F}_k$ has an orthogonal complement $Q_k \in \mathbb{R}^{D\times(D-m)}$ such that, $Q_k^T \bar{F}_k = 0$. Then the geodesic flow, $\phi(t) : t \in [0, 1]$, between $\bar{F}_k$ and $\bar{F}_{k+1}$ is such that, $\phi(t) \in \mathbb{G}_{m,D}$ and $\phi(0) = \bar{F}_k$ and $\phi(1) = \bar{F}_{k+1}$. This implies that starting from $\bar{F}_k$, the geodesic flow reaches $\bar{F}_{k+1}$ in unit time and its expression is given by

$$\phi(t) = \bar{F}_k \Delta_1 \Lambda(t) - Q_k \Delta_2 \Omega(t) \tag{1}$$

where $\Delta_1 \in \mathbb{R}^{m\times m}$ and $\Delta_2 \in \mathbb{R}^{(D-m)\times m}$ are orthonormal matrices. $\Delta_1$ and $\Delta_2$ can be obtained using the following equations

$$\begin{aligned} \bar{F}_k^{'} \bar{F}_{k+1} &= \Delta_1 \Lambda V^{'} \quad (2)\\ Q_k^{'} \bar{F}_{k+1} &= -\Delta_2 \Omega V^{'} \quad (3) \end{aligned}$$

where $\Lambda, \Omega \in \mathbb{R}^{m\times m}$ are diagonal matrices whose diagonal elements are $cos\theta_j$ and $sin\theta_j$ for $j = 1, 2, \ldots m$. $\{\}^{'}$ denotes the transpose operator. $\theta_j$ are known as the principal angles between $\bar{F}_k$ and $\bar{F}_{k+1}$. By using different values of $t$, we can obtain different intermediate subspaces.

### 2.2. Computation of aligning matrices

In the next step, all the virtual intermediate subspaces are aligned with the frontal subspace using pose-specific alignment matrices. The reason for alignment of subspaces is two-fold: (1) Matching performance is considerably better if both the gallery and probe images are in the frontal pose as compared to both being in the same, non-frontal pose. To illustrate this, we have taken 100 subjects from the Multi-PIE dataset [13] having frontal pose and frontal illumination condition. They are down-sampled and then up-sampled to get their low resolution versions. Nearest neighbour search with the high-resolution (HR) images as gallery and low-resolution (LR) images as probe gave recognition rate of 80%. In contrast, if we perform the same experiment for pose 04_1 (both gallery and probe have the same non-frontal pose), the performance drops to 67%, which justifies our assumption. (2) Compared to subspace to point representation as used in [12], this approach results in much lower dimension of the descriptor, lower computational time as well as better recognition performance as shown later.

Now we describe in detail how we perform the alignment (Fig. 1). Suppose $\bar{F}_f$ denote the frontal subspace and $\bar{F}_{g_1}, \bar{F}_{g_2}, \ldots, \in \mathbb{R}^{D\times m}$ denote the generated intermediate subspaces (non-frontal). We compute the frontal alignment matrices $A_i$ for each of these intermediate generated subspaces and the frontal subspace by minimizing the Bregman matrix divergence as follows

$$A_i^* = argmin_{A_i} \left\| \bar{F}_{g_i} A_i - \bar{F}_f \right\|_F \tag{4}$$

where, $\|.\|_F$ is the Frobenius norm. Since $\bar{F}_f$ and $\bar{F}_{g_i}$ are generated from the first $m$ normalized eigenvectors of the corresponding subspaces, they are intrinsically regularized and adding a regularizer term is not required in (4) [16]. A closed form solution can be obtained for (4) as $A_i^* = \bar{F}_{g_i}^{'} \bar{F}_f$. Thus, if the total number of subspaces is $N$, there are $N-1$ alignment matrices, $A_1^*, A_2^*, \ldots, A_{N-1}^* \in \mathbb{R}^{m\times m}$. Then, each non-frontal subspace is aligned as follows

$$\bar{F}_{g_i}^* = \bar{F}_{g_i} A_i^* \tag{5}$$

where, $\bar{F}_{g_i}^*$ is the aligned subspace.

### 2.3. Computation of discriminative features

After generating the intermediate subspaces and aligning them, we project the low level feature $p$ computed from an image to all the aligned subspaces including the frontal one. Next, we utilize the class labels of the training data to learn a Mahalanobis distance metric $T$ to make the features discriminative. Instead of learning only one metric for the entire data, we divide the aligned features from different subspaces into few sets and learn a metric [17] for each set.

We consider features from the same subject/class having variations in pose, jointly pose and resolution as the match pairs and those from different subjects/classes as the non-match pairs. We decide on whether any given pair of features $p_i$ and $p_j$, belong to same class or not, from likelihood ratio test as formulated below

$$\psi(p_i, p_j) = log\left(\frac{prob(p_i, p_j|\Theta_0)}{prob(p_i, p_j|\Theta_1)}\right) \tag{6}$$

where, $\Theta_0$ and $\Theta_1$ are the hypotheses that a pair is non-match and match respectively. The value of $\psi(p_i, p_j)$ is small when a pair of features belong to the same class and its value is large when a pair of features belong to different class. Assuming a Gaussian structure for the difference space of features, and plugging them into (6), we get the simplified equation

$$\psi(p_{ij}) = p_{ij}^{'}\left(C_{n_{ij}=1}^{-1} - C_{n_{ij}=0}^{-1}\right) p_{ij} \tag{7}$$

where, $p_{ij} = p_i - p_j$ is a vector in the difference space; $n_{ij} = 1$ for a matched pair and its value is 0 for a non-matched pair. $C_{n_{ij}=1}$ and $C_{n_{ij}=0}$ are the corresponding covariance matrices. The final transformation is given by $T = \left(C_{n_{ij}=1}^{-1} - C_{n_{ij}=0}^{-1}\right)$. $T$ is made positive semidefinite by clipping the spectrum of $T$ by eigen analysis.

During testing, the low-level image feature is projected onto all the aligned subspaces and then transformed using the discriminative transform learned in the training phase. Finally they are concatenated to get the *ADPR* descriptor.

## 3. EXPERIMENTAL EVALUATION

In this section, we present the results of extensive experiments conducted on face recognition across pose and resolution, and object recognition across pose to test the applicability of the proposed approach for these applications.

### 3.1. Face Recognition Across Pose and Resolution

Here, we test the applicability of the proposed descriptor for recognizing faces across multiple variations. Specifically, we perform face recognition with frontal and high resolution (HR) images as gallery and non-frontal, low-resolution (LR) images under varying illuminations as probe, as usually found in surveillance scenarios. Face images are represented by local feature descriptors (SIFT [18] in this paper) computed at 15 fudicial locations.

**Results on MultiPIE dataset:** First we report results on the Multi-PIE dataset [13] containing images of 337 subjects from four different recording sessions captured under different poses, illumination conditions and expressions. We follow the same experimental protocol as [12]. We generate 8 intermediate subspaces between HR frontal and LR (down-sampled by a scale of three) pose 04_1 and 13_0 during training. Results for the proposed descriptor and comparisons with several state-of-the-art approaches are reported in Table 1. These results are either directly taken from [12] or generated from the codes provided by the authors (for the more recent papers [19], [20]). Even though the two intermediate poses (14_0 and 05_0) are not used during training, we still achieve good performance for probe images in these poses, whereas results for all the other approaches reported in Table 1 (except [12]) are obtained using all the poses for training. The performance for these approaches are considerably lower when they are trained with only the frontal and extreme poses, as used in the proposed approach and [12].

**Table 1**. Rank-1 recognition performance (%) for four different probe poses, averaged over the different gallery illuminations on the Multi-PIE dataset [13].

| Method | Pose 13_0 | Pose 14_0 | Pose 05_0 | Pose 04_1 |
|---|---|---|---|---|
| MDS Learning [21] | 32.8 | 44.8 | 47.0 | 48.5 |
| LSML [17] | 46.9 | 53.9 | 55.2 | 54.3 |
| GMA [22] | 65.0 | 70.1 | 70.3 | 64.2 |
| MvDA [19] | 45.7 | 55.0 | 53.8 | 42.9 |
| FCPRF + LSML [20] | 54.0 | 71.2 | 73.4 | 61.0 |
| SCDL [23] | 66.3 | 73.0 | 72.7 | 64.1 |
| CFDL [24] | 65.9 | 72.0 | 72.8 | 64.7 |
| SCDL + LSML | 69.1 | 75.1 | 74 | 67.6 |
| CFDL + LSML | 68.9 | 74.1 | 74.6 | 68.1 |
| DPFD [12] | 74.5 | 78.0 | 74.0 | 70.1 |
| **Proposed *ADPR*** | **75.3** | **78.0** | **76.1** | **72.0** |

**Results on Surveillance Cameras Face Database:** We also evaluate the proposed descriptor on real surveillance quality data obtained from the Surveillance Cameras Face (SCface) Database [10]. This data contains images of 130 subjects captured in uncontrolled environment using five different video surveillance cameras. Following the experimental protocol of [12], we randomly select 50 subjects for training and the remaining 80 for testing. During training, we generate 4 intermediate virtual subspaces in between the HR frontal and LR non-frontal images from one camera. For all the other approaches, we use two setups for training: (1) similar setup as our proposed approach, (2) HR frontal and LR images from all the five cameras. This experiment shows that *ADPR* can generalize better across unseen poses. Even though only one camera is used for training, *ADPR* performs better significantly than the other approaches.

### 3.2. Object Recognition Across Pose

Here we evaluate the applicability of the proposed descriptor to recognize general objects on two datasets, namely the Columbia Object Image Library 20 (COIL 20) database [11] and the RGB-D Object Database [14].

**Table 2**. Rank-1 accuracy (%) of the proposed approach and comparison with state-of-the-art approaches on the Surveillance Cameras Face Database [10].

| Method | Rank-1 | Rank-1 |
|---|---|---|
| | 1 Cam | 5 Cam |
| MDS Learning [21] | 30.0 | 61.1 |
| LSML [17] | 64.7 | 67.2 |
| GMA [22] | 38.2 | 50.5 |
| FCPRF + LSML [20] | 58.0 | 61.3 |
| SCDL [23] | 48.2 | 58.5 |
| CFDL [24] | 45.7 | 62.2 |
| SCDL + LSML | 48.8 | 60.0 |
| CFDL + LSML | 46.3 | 63.3 |
| DPFD [12] | 69.0 | – |
| **Proposed *ADPR*** | **73.3** | – |

**Results on COIL 20 database:** For the COIL 20 database [11], 50 images of each object that has pose variations from left extreme to right extreme including the frontal pose are selected for the experiments. We follow the same experimental protocol as in [12]. We have resized the images to 32 × 32 and used the image intensity values as the input features. The results of the proposed descriptor and comparisons with the state-of-the-art are given in Table 3.

**Table 3**. Rank-1 accuracy (%) of the proposed approach and comparison with other approaches on COIL 20 Database [11].

| Method | Rank-1 Accuracy |
|---|---|
| MDS Learning [21] | 75.6 |
| LSML [17] | 80.3 |
| GMA [22] | 66.1 |
| SCDL [23] | 79.2 |
| CFDL [24] | 78.7 |
| SCDL + LSML | 82.6 |
| CFDL + LSML | 82.0 |
| MvDA [19] | 69.7 |
| DPFD [12] | 82.2 |
| **Proposed *ADPR*** | **83.0** |

**Results on RGB-D Object Database:** We have also performed experiments on a larger object database (RGB-D Object Database [14]) which contains both RGB and depth images from 51 categories. The objects are captured in such a way that they are covered from multiple views. We have taken all the images (both visual and depth) of the first instance in each category for our experiments. In each category, we select five images from four different poses for training and use the rest of the images as probe during testing.

We have extracted kernel descriptors [25] of dimension 500 separately from visual and depth images to use as features in this experiment. Table 4 shows the recognition performance for recognizing visual probe images against visual gallery images (Visual - Visual) and depth probe images against depth gallery images (Depth - Depth). Comparison with other algorithms is also shown for both the cases. We observe that the proposed *ADPR* descriptor has an excellent performance as compared to the other approaches thus justifying its usefulness for the application of general object recognition.

### 3.3. Analysis with state-of-the-art deep features

Here, we analyze the proposed approach with some of the recent deep learning methods to show that the proposed descriptor can be used to further boost the performance of features obtained from deep neural networks. For this purpose, we perform experiments on the Multi-PIE dataset [13] for faces.The experimental setup is the same as in Section 3.1. For the Multi-PIE dataset [13], we use the output of FC6 layer of a recent deep learning architecture VGGNet [26], as features in this experiment. Using the HR images as gallery and the LR images as probe and nearest neighbour classifier, the rank 1

**Table 4**. Rank-1 accuracy (%) of the proposed approach and comparison with other approaches on RGB-D Object Database [14].

| Method | Visual - Visual | Depth - Depth |
|---|---|---|
| MDS Learning [21] | 82.2 | 53.9 |
| LSML [17] | 60.1 | 45.8 |
| GMA [22] | 70.6 | 38.9 |
| MvDA [19] | 77.2 | 50.6 |
| SCDL [23] | 80.4 | 61.1 |
| CFDL [24] | 81.0 | 60.5 |
| SCDL+LSML | 81.7 | 62.0 |
| CFDL+LSML | 82.0 | 61.3 |
| DPFD [12] | 86.0 | 62.0 |
| **Proposed *ADPR*** | **91.4** | **69.2** |

**Table 5**. Rank-1 recognition accuracy (%) for four different probe poses, averaged over the different gallery illuminations on the Multi-PIE dataset [13] using VGG Features [26].

| Method | Pose 13_0 | Pose 14_0 | Pose 05_0 | Pose 04_1 |
|---|---|---|---|---|
| VGG-HR-LR-NN | 32.2 | 52.8 | 53.1 | 32.8 |
| **VGG-HR-LR-ADPR** | **39.6** | **54.6** | **55.3** | **39.2** |
| VGG-HR-HR-NN | 88.3 | 97.0 | 97.0 | 91.3 |
| **VGG-HR-HR-ADPR** | **91.9** | **98.0** | **98.0** | **93.9** |

accuracy (%) is reported in Table 5, denoted as VGG-HR-LR-NN. Since the VGGNet is not trained on low resolution images, the performance is quite low as expected. The performance of the proposed *ADPR* descriptor using the VGGNet [26] output as the low-level features is also reported as VGG-HR-LR-ADPR. We see that though the performance is still low, it is better as compared to using the VGGNet features directly. Since the VGGNet is trained on HR images, we also perform another experiment with both HR images as gallery and probe. We observe that the proposed *ADPR* descriptor is able to further improve the performance thus justifying its usefulness with different kinds of input features.

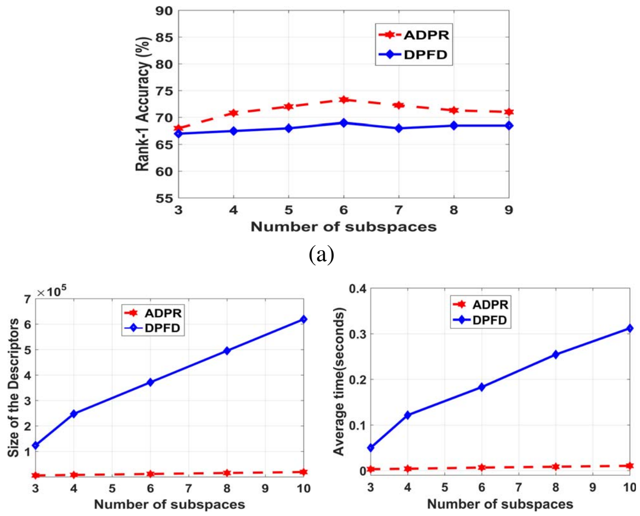
For the application of object recognition, we perform similar experiments on RGB-D database [14] with AlexNet [27], pretrained on object images from ImageNet. The experimental setup is the same as described in Section 3.2. We obtain an accuracy of 90.2% when we take the output of the first fully connected layer and use nearest neighbour classifier. The results using the AlexNet along with the proposed descriptor is 93.4%. Note that both these deep networks have been trained on millions of images, and so improvement over these features using the proposed descriptor justifies the usefulness of the proposed approach.

### 3.4. Analysis of proposed descriptor

Here we analyze and compare the proposed descriptor, *ADPR* with *DPFD* in more detail since they are most appropriate for pose-free applications. For this purpose, we have chosen the Surveillance Cameras Face Database [10]. The experimental setup is similar to that of [12].

First, we analyze the effect of the number of subspaces on the Rank-1 accuracy. This result is shown in Figure 2(a).We observe that the performance of the two descriptors does not vary widely but the performance using *ADPR* is always superior than *DPFD* for a wide range of the number of subspaces.

Now, we analyse the feature dimension and the computational requirements of the two descriptors.The dimensions of the two descriptors, *ADPR* and *DPFD* are functions of the number of intermediate subspaces used. Figure 2(b) shows the variation in the feature dimension of *ADPR* and *DPFD* with different number of intermediate subspaces. For this dataset, we have taken the number of subspaces as six. Therefore the feature dimensions for *DPFD* and *ADPR* descriptors are 371520 and 11520 respectively where each facial image is represented as concatenation of features computed from 15 fiducial points on the face. We observe that the dimension of both the descriptors increases with the number of subspaces. But the feature dimension of *ADPR* is much less as compared to that of *DPFD* descriptor for the entire range.



(a)

(b) (c)

**Fig. 2**. Comparison curves. (a) Rank-1 accuracy (%) vs number of subspace (b) Descriptor size vs number of subspace (c) Time (seconds) vs number of subspace.

Since the time required to compute the distance between two descriptors is a function of their dimension and we have already observed that the dimension of *ADPR* is considerably less than that of *DPFD*, it is expected that it will take less time to compute the distance between two *ADPR* descriptors as compared to two *DPFD* descriptors. Figure 2(c) shows the plot of time required for pairwise comparison (in seconds) against the number of subspaces. Here also, we observe that the time required for *ADPR* is much less than that required for *DPFD*. For the SC-Face [10] Database, there are 80 gallery images during testing, so the time required to get the identity of one probe image is around 0.54 sec for *ADPR*, in comparison to around 15 sec for *DPFD*. This difference will increase as the size of the gallery increases.

## 4. CONCLUSION

In this work, we proposed a novel aligned discriminative pose robust descriptor (*ADPR*) for matching faces and objects across pose. It requires images from a few regions of the pose space for training and does not require separate training for each probe pose. Experimental evaluation and analysis prove the usefulness and generalizability of the proposed approach for the task of face and object recognition across a wide range of variations like pose, illumination and resolution.

## 5. REFERENCES

[1] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Transactions on Pattern Analysis and machine Intelligence*, vol. 38, no. 3, pp. 518 – 531, 2016.

[2] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face

recognition," *IEEE Transactions on image Processing*, vol. 24, no. 3, pp. 980–993, 2015.

[3] J. Gu, H. Hu, H. Li, and W. Hu, "Patch-based alignment-free generic sparse representation for pose-robust face recognition," *ICIP*, pp. 3006–3010, 2016.

[4] X. Duan and Z. H. Tan, "Local feature learning for face recognition under varying poses," *ICIP*, pp. 2905–2909, 2015.

[5] C. Ren, D. Dai, K. Huang, and Z. Lai, "Transfer learning of structured representation for face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5440–5454, 2014.

[6] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha, "Improving cross-resolution face matching using ensemble-based co-transfer learning," *IEEE Transactions on image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CVPR*, pp. 815–823, 2015.

[8] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *CVPR*, pp. 1701–1708, 2014.

[9] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," *arXiv preprint arXiv:1601.04153*, 2016.

[10] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.

[11] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-20)," *Technical Report*, 1996.

[12] S. Sanyal, S. P. Mudunuri, and S. Biswas, "Discriminative pose-free descriptors for face and object matching," *ICCV*, pp. 3837–3845, 2015.

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

[14] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," *ICRA*, pp. 1817–1824, 2011.

[15] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2288–2302, 2014.

[16] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," *CVPR*, pp. 2960–2967, 2013.

[17] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," *CVPR*, pp. 2228–2295, 2012.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60(2), pp. 91–110, 2004.

[19] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis and machine Intelligence*, vol. 38, no. 1, pp. 188–194, 2016.

[20] F. Shen, C. Shen, X. Zhou, Y. Yang, and H. T. Shen, "Face image classification by pooling raw features," *Pattern Recognition*, vol. 54, pp. 94–103, 2016.

[21] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer, "Pose-robust recognition of low-resolution face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3037–3049, 2013.

[22] A. Sharma, A. Kumar, H. Daume, and D.H. Jacobs, "Generalized multiview analysis: A discriminative latent space," *ICCV*, pp. 2160–2167, 2012.

[23] S. Wang, D. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," *CVPR*, pp. 2216–2223, 2012.

[24] D. A. Huang and Y. C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," *ICCV*, pp. 2496–2503, 2013.

[25] L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," *In Advances in Neural Information Processing Systems*, pp. 244–252, 2010.

[26] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," *BMVC*, pp. 1–6, 2015.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, pp. 1097–1105, 2012.