Aligned Discriminative Pose Robust Descriptors for Face and Object Recognition

Soubhik Sanyal, Devraj Mandal, Soma Biswas

Image Analysis and Computer Vision Lab
Department of Electrical Engineering
Indian Institute of Science, Bangalore, India

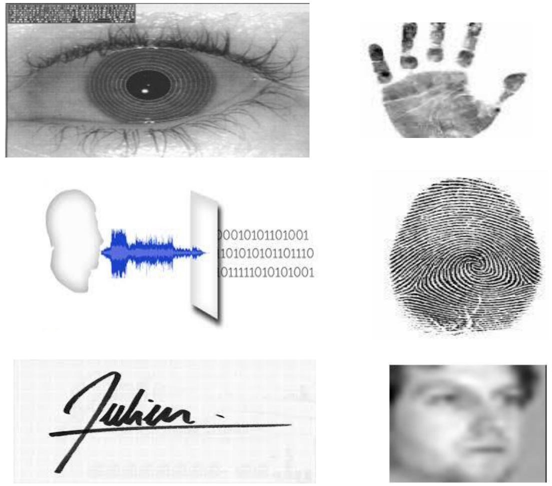
{soubhiksanyal, devraj89, soma.biswas}.ee.iisc.ernet.in

Date 19/09/2017

Outline:

- Introduction and Problem Statement
- Related work and Limitations
- Proposed Framework
- Experimental Results
- Analysis of the Descriptors
- Conclusion
- References

Introduction



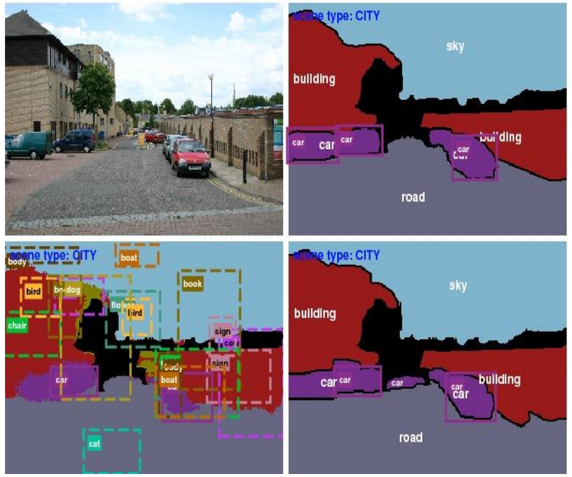
Biometrics

Tagging friends

Online Payments

Catching Criminals!

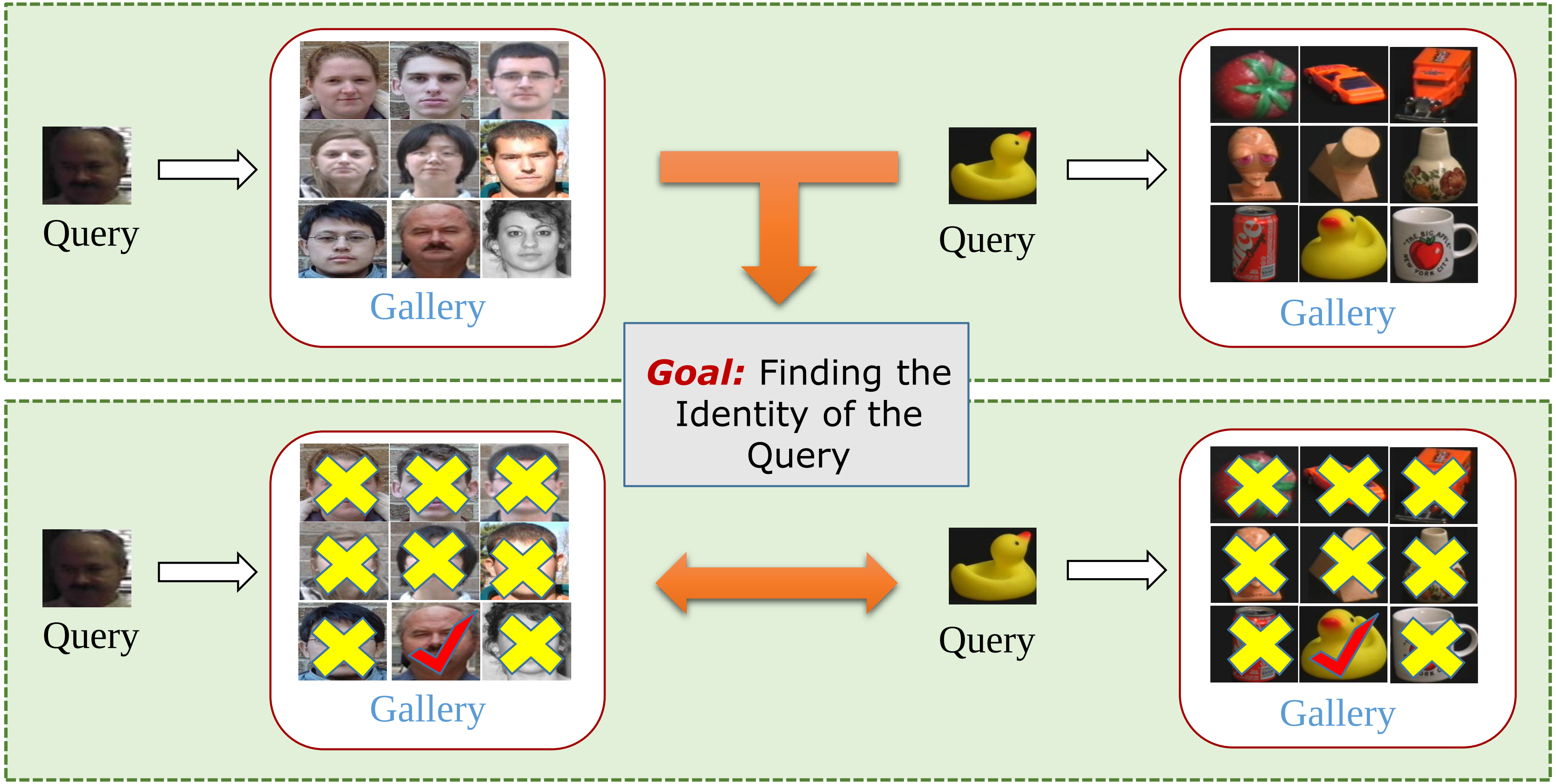
Autonomous Cars



Scene Understanding

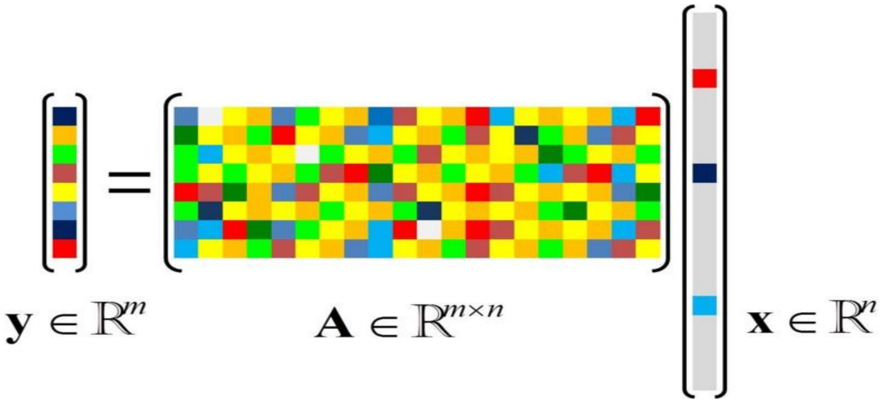
(Images are taken from Google)

Problem Statement

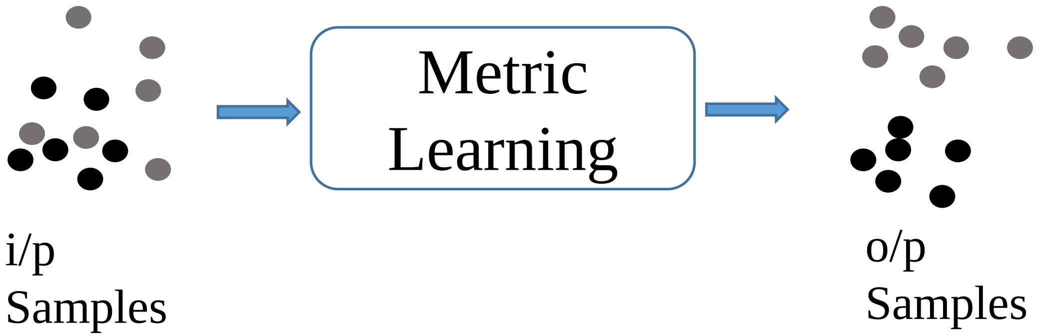


(Images are taken from MBGC dataset and COIL 100 dataset)

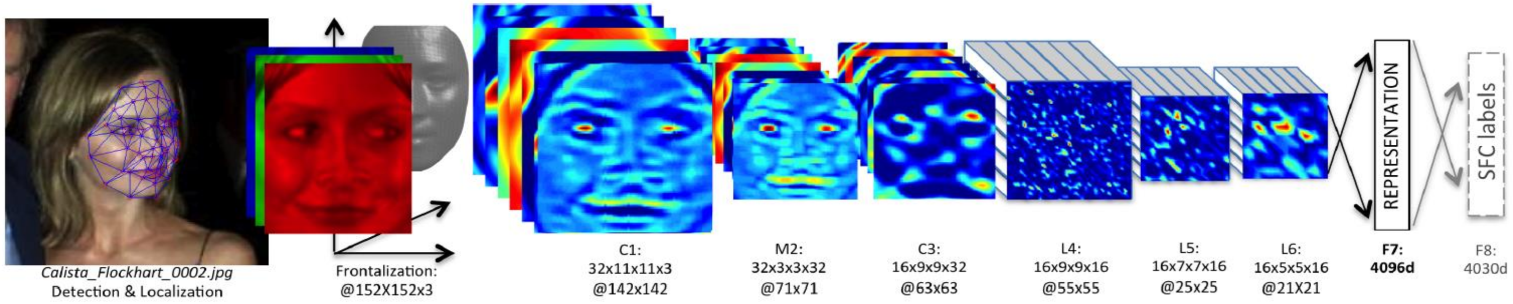
Related Work



Dictionary Learning [CVPR 2012, ICCV 2013, PAMI 2016]



Metric Learning [CVPR 2012, PAMI 2016]



Convolutional Neural Networks [CVPR 2014, BMVC 2015]

(Images are taken from Google)

Limitations of the Existing Approaches

- Required samples from every available pose

- Poor performance for unseen poses

Proposed Framework: Computation of low level features

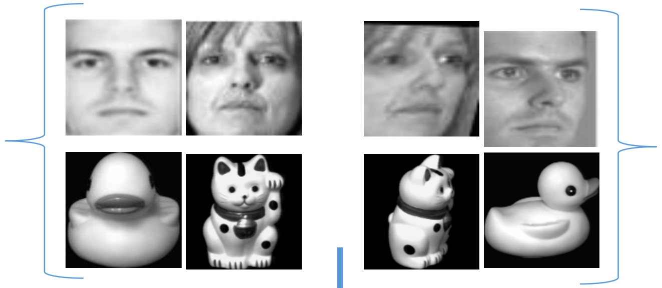
Training data
at few poses

Generate features

SIFT Features [D. G. Lowe, IJCV 2004] at fiducial location

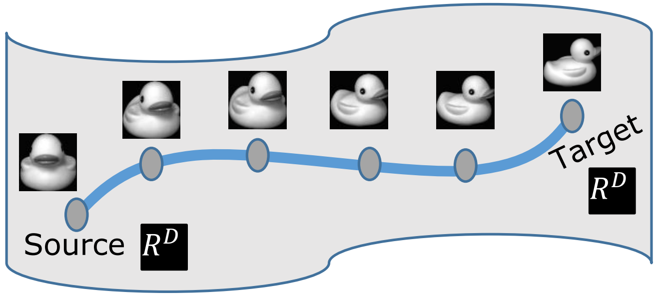
Intensity features, Kernel descriptors

Proposed Framework: Generating synthetic pose subspaces



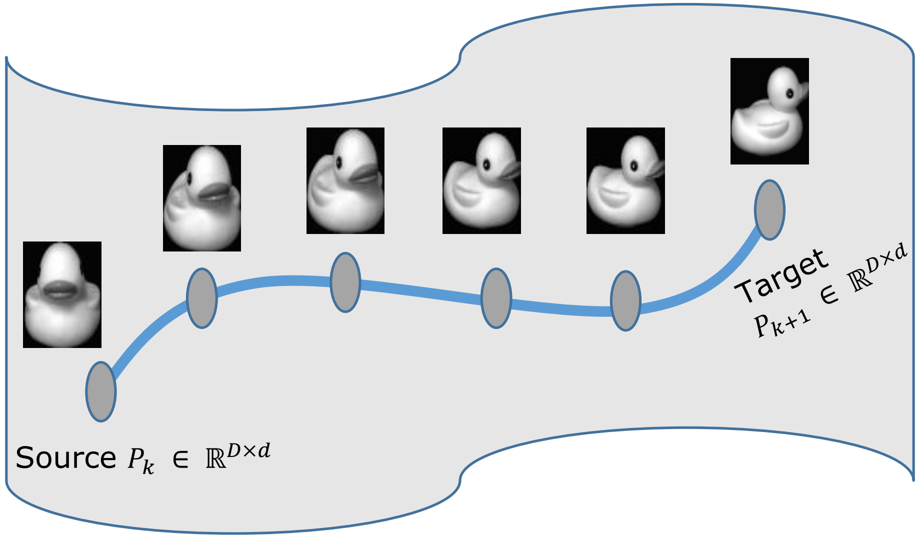
Training data
at few poses

Generate features



Generate intermediate subspaces by sampling the geodesic flow path

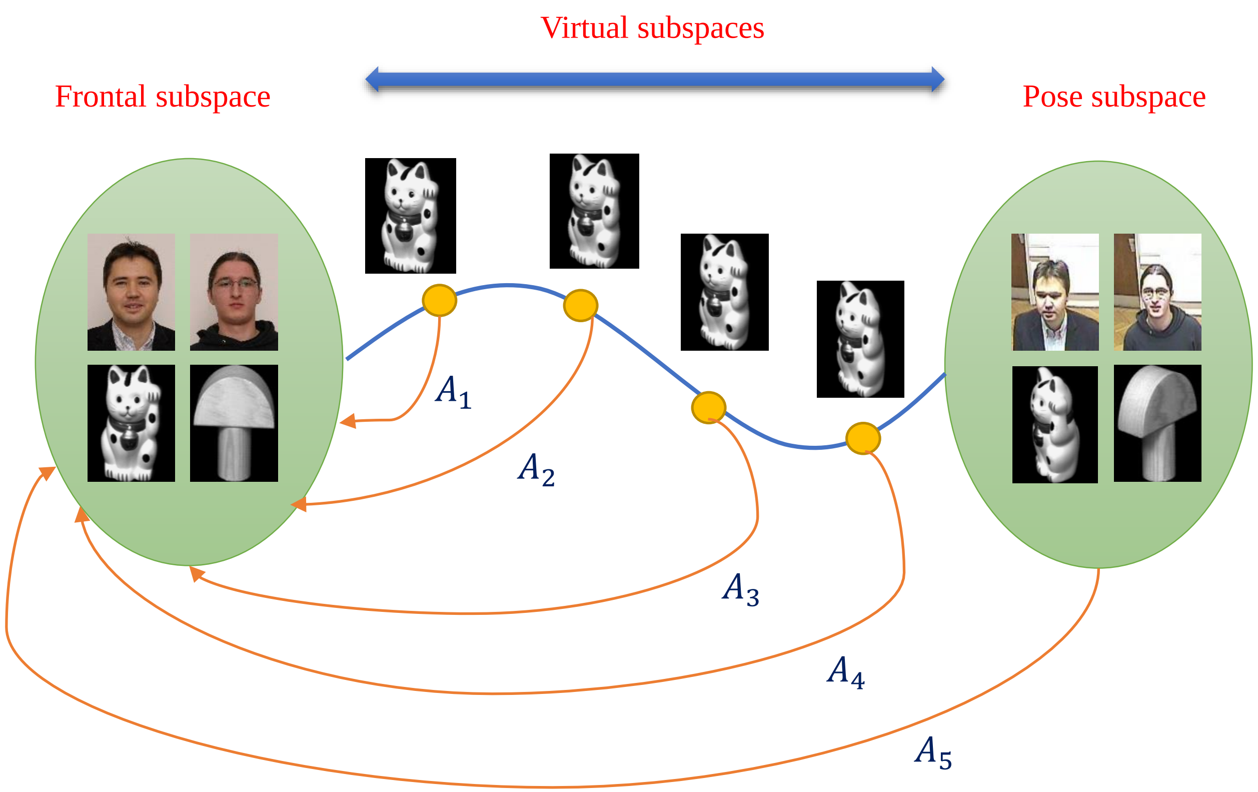
Proposed Framework: Generating synthetic pose subspaces



- Source Subspace: $P_k \in \mathbb{R}^{D\times d}$ and target subspace: $P_{k+1} \in \mathbb{R}^{D\times d}$
- Geodesic Flow: $\psi(t) = P_k U_1 \Gamma(t) - R_k U_2 \Sigma(t)$
- Orthogonal Complement of $P_k \Rightarrow R_k \in \mathbb{R}^{D\times(D-d)}$ such that $R_k^T P_k = 0$
- $t \in [0,1]$ and $\psi(0) = P_k$ and $\psi(1) = P_{k+1}$
- $P_k^T P_{k+1} = U_1 \Gamma V^T$
- $R_k^T P_{k+1} = -U_2 \Sigma V^T$
- $\Gamma, \Sigma \in \mathbb{R}^{d\times d}$ diagonal matrices with diagonal elements as $\cos\theta_i, \sin\theta_i$ for $i = 1,2,\ldots\ldots,d$.
- θ_i: principal angles between P_k and P_{k+1}

Reference: Gopalan, Li and Chellappa, PAMI 2014

Proposed Framework: Flow Chart of computing ADPR Descriptor



Aligning Every Subspace with the Frontal one

Proposed Framework: Motivation for ADPR

- ❑ Computational time to recognize a subject is higher for DPFD [ICCV 2015] if the number of subspaces increased.

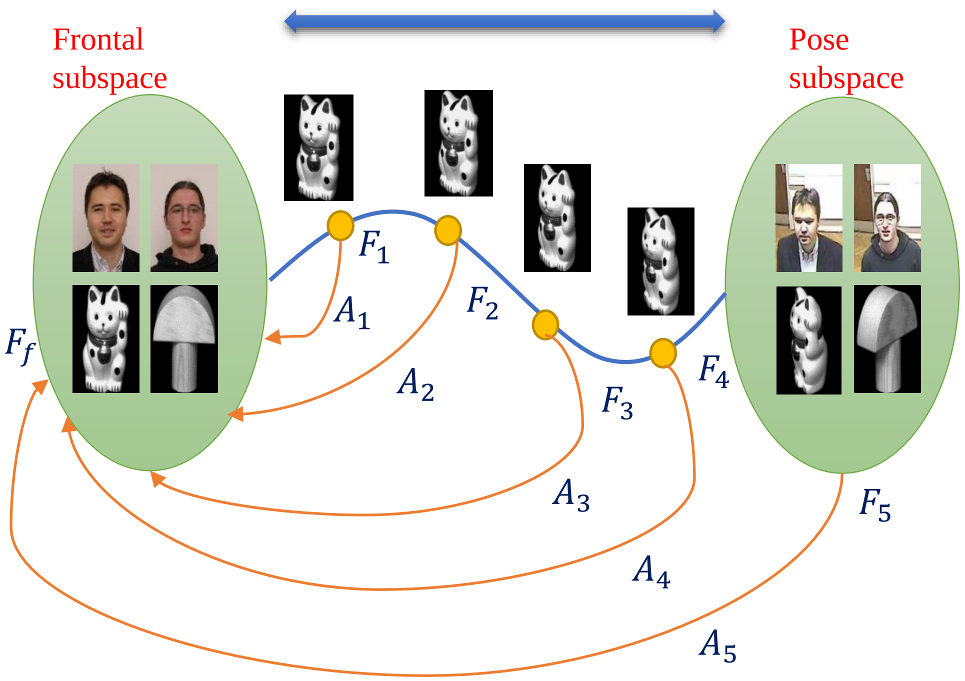
- ❑ Matching performance is considerably better if gallery and probe both are in frontal pose as compared to both being in same non-frontal pose.

- ❑ Experiment: 100 subjects from Multi-PIE [R. Gross *et al.* 2007] dataset with frontal pose and illumination.

- ❑ Down-sampled and up-sampled to get low resolution version: Rank-1 recognition rate = 80%.

- ❑ Similar experiment with Pose04_1 (30 degree): Rank-1 recognition rate = 67%.

Proposed Framework: Computation of Aligning Matrices

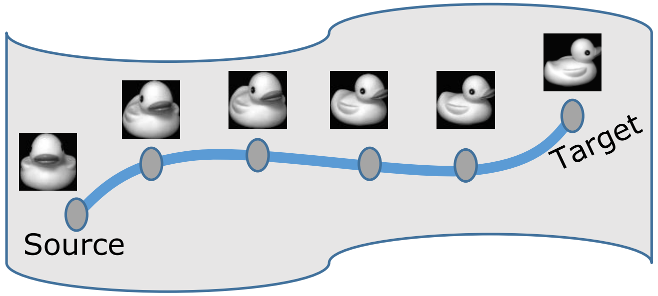


Aligning Every Subspace with the Frontal one

- $A_i^* = argmin_{A_i} \|F_i A_i - F_f\|_F^2$
- F_i: i^{th} subspace, A_i: i^{th} alignment matrix, F_f : frontal subspace
- $A_i^* = argmin_{A_i} \|F_i^T F_i A_i - F_i^T F_f\|_F^2$
- $A_i^* = argmin_{A_i} \|A_i - F_i^T F_f\|_F^2$
- $A_i^* = F_i^T F_f$
- h subspaces $\Longrightarrow$ $(h-1)$ number of alignment matrices
- Size of ADPR descriptor with six subspaces: $128 \times 6 = 768$
- Size of DPFD descriptor with six subspaces: 24768 [ICCV 2015]

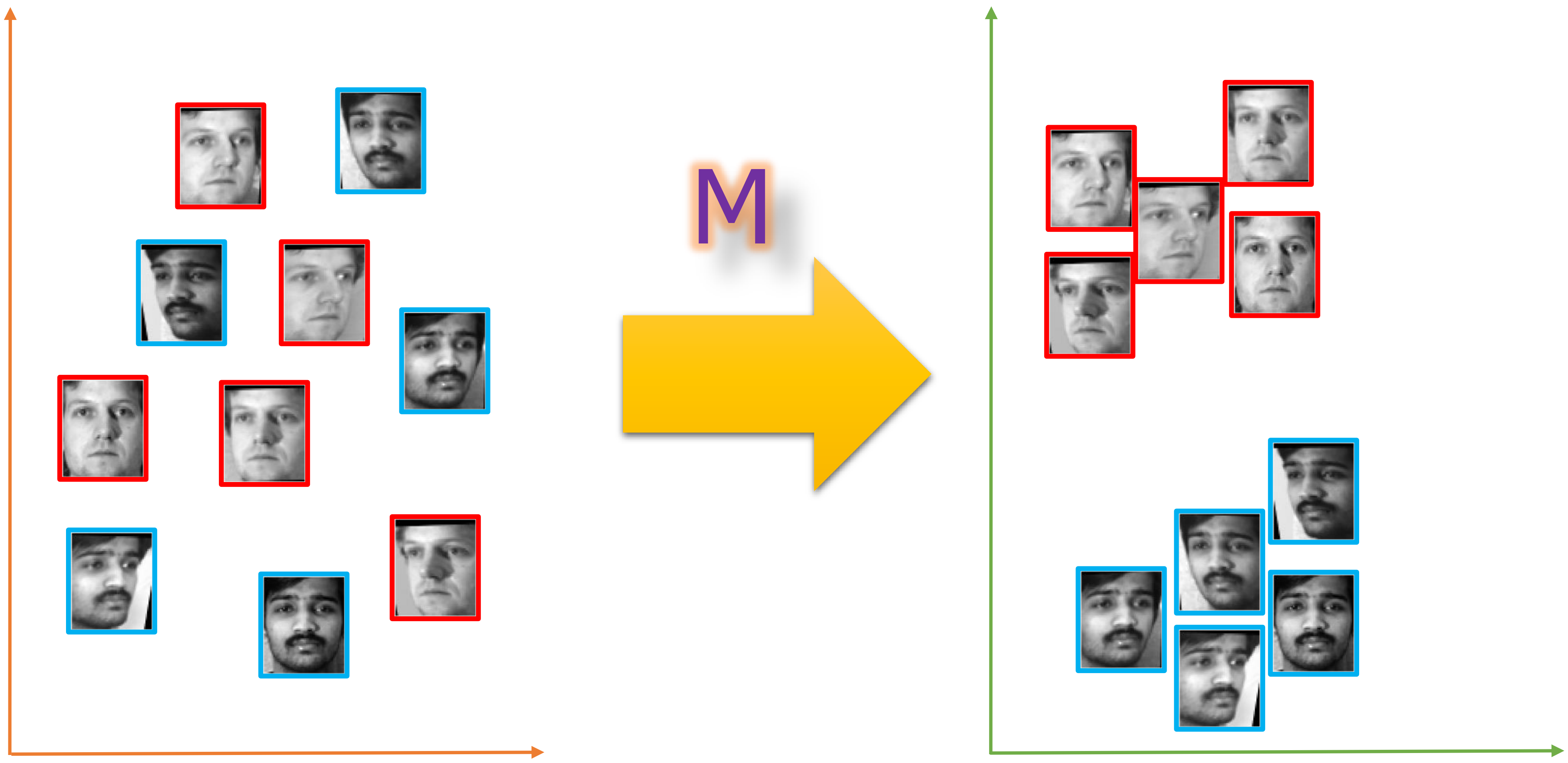
Proposed Framework: Projection into the intermediate subspaces

Training data
at few poses



Generate features → Generate intermediate subspaces by sampling the geodesic flow path → Project training data onto Aligned Subspaces

Proposed Framework: Making the features discriminative



(Images are taken from Multi-PIE dataset [R. Gross *et al.* 2007])

Proposed Framework: Making the features discriminative

Matched Pairs

Non-matched Pairs

- Mahalanobis metric: $d^2(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j)$

- Decision on matched or non-matched pairs depending upon likelihood ratio test

$$\delta(x_i, x_j) = log\left(\frac{p(x_i, x_j | H_0)}{p(x_i, x_j | H_1)}\right)$$

- H_0 : Hypothesis that the pair is non-matched, H_1 : Hypothesis that the pair is matched

Proposed Framework: Making the features discriminative

- Probabilities in difference space

$$p(x_i, x_j | H_0) = \frac{1}{\sqrt{2\pi \left|\Sigma_{n_{ij}=0}\right|}} e^{\left(-\frac{1}{2} x_{ij}^T \Sigma_{n_{ij}=0}^{-1} x_{ij}\right)}$$

$$p(x_i, x_j | H_1) = \frac{1}{\sqrt{2\pi \left|\Sigma_{n_{ij}=1}\right|}} e^{\left(-\frac{1}{2} x_{ij}^T \Sigma_{n_{ij}=1}^{-1} x_{ij}\right)}$$

$$\Sigma_{n_{ij}=1} = Covariance\ matrix\ of\ matched\ pairs$$

$$\Sigma_{n_{ij}=0} = Covariance\ matrix\ of\ nonmatched\ pairs$$

$$x_{ij} = x_i - x_j = Difference\ Space$$

Proposed Framework: Making the features discriminative

- After putting the probabilities and simplification

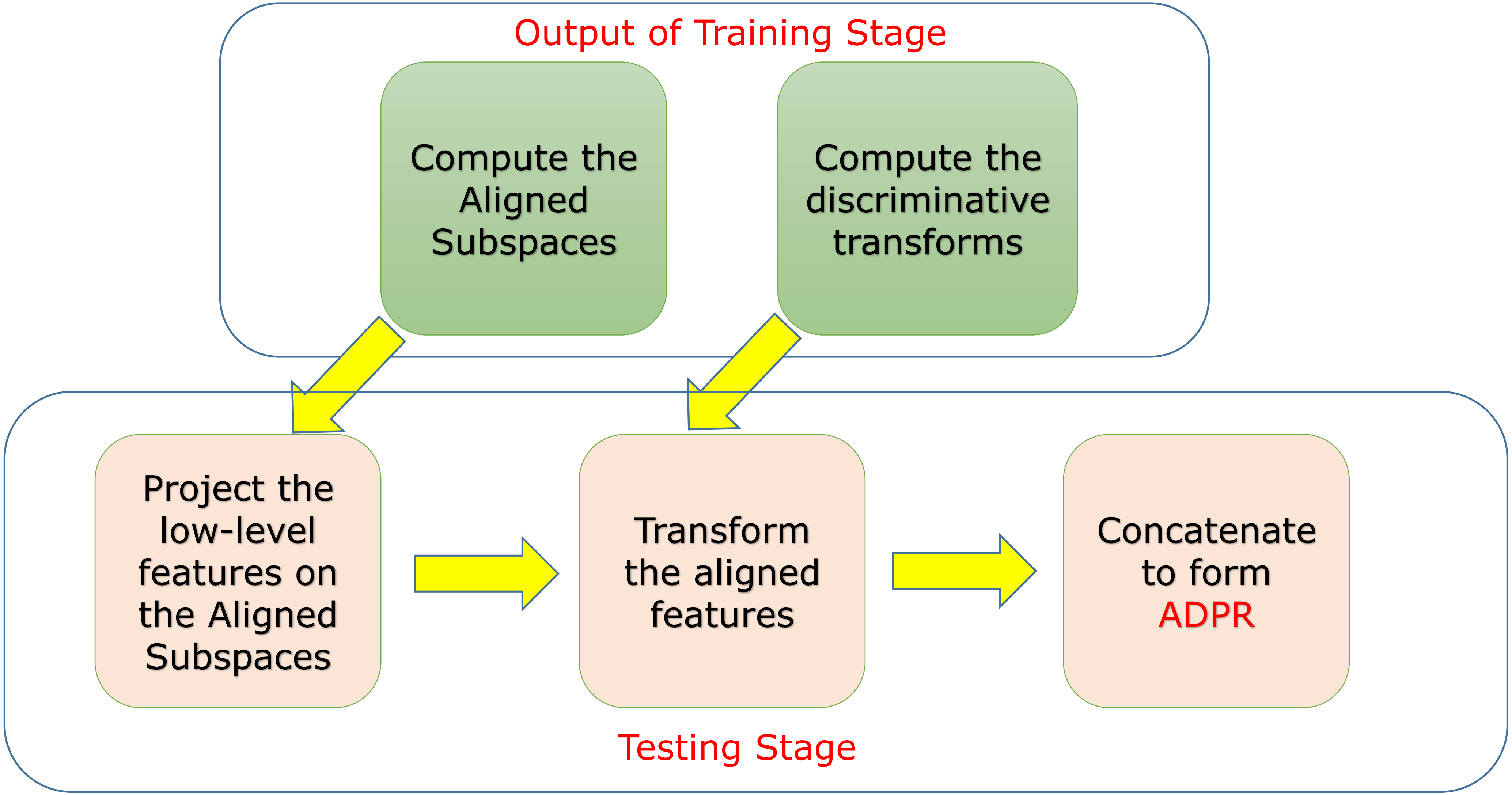
$$\delta(x_{ij}) = \log\left(\frac{\frac{1}{\sqrt{2\pi\left|\Sigma_{n_{ij}=0}\right|}}e^{\left(-\frac{1}{2}x_{ij}^{T}\Sigma_{n_{ij}=0}^{-1}x_{ij}\right)}}{\frac{1}{\sqrt{2\pi\left|\Sigma_{n_{ij}=1}\right|}}e^{\left(-\frac{1}{2}x_{ij}^{T}\Sigma_{n_{ij}=1}^{-1}x_{ij}\right)}}\right)$$

$$\delta(x_{ij}) = x_{ij}^{T}\left(\Sigma_{n_{ij}=1}^{-1} - \Sigma_{n_{ij}=0}^{-1}\right)x_{ij}$$

- Analysing

$$\mathrm{M} = \left(\Sigma_{n_{ij}=1}^{-1} - \Sigma_{n_{ij}=0}^{-1}\right)$$

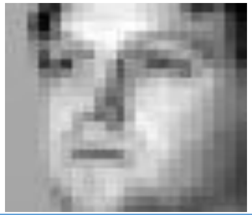
Proposed Framework: Computation for ADPR

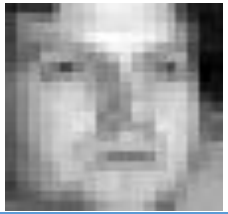


Experimental Analysis:

- MultiPIE Dataset [R. Gross *et al.* 2007]: 200 subjects in 5 poses and 20 illumination conditions
- SC Face Dataset [M. Grgic *et al.* 2011]: 130 subjects captured in real surveillance cameras in 5 different poses
- Coil-20 Dataset [S.A. Nene *et al.* 1996]: 20 subjects in multiple views
- RGB-D Dataset [K. Lai *et al.* 2011]: 51 subjects in multiple views

Experiments on Face Recognition Across Pose and Resolution






Method	Pose 13_0	Pose 14_0	Pose 05_0	Pose 04_1
MDS Learning [S. Biswas *et al.* TPAMI 2013]	32.8	44.8	47.0	48.5
LSML [M. Kostinger *et al.* CVPR 2012]	46.9	53.9	55.2	54.3
GMA [A. Sharma *et al.* ICCV 2012]	65.0	70.1	70.3	64.2
MvDA [M. Kan *et al* PAMI 2016]	45.7	55.0	53.8	42.9
FCPRF + LSML [F. Shen *et al.* PR 2016]	54.0	71.2	73.4	61.0
SCDL [S. Wang *et al.* CVPR 2012]	66.3	73.0	72.7	64.1
SCDL + LSML	69.1	75.1	74	67.6
CFDL [D.A. Huang *et al.* ICCV 2013]	65.9	72.0	72.8	64.7
CFDL + LSML	68.9	74.1	74.6	68.1
DPFD [ICCV 2015]	74.5	78.0	74.0	70.1
Proposed (ADPR)	**75.3**	**78.0**	**76.1**	**72.0**

Rank-1 recognition performance (%) for four different probe poses, averaged over the different gallery illuminations on the Multi-PIE dataset [R. Gross *et al.* 2007]

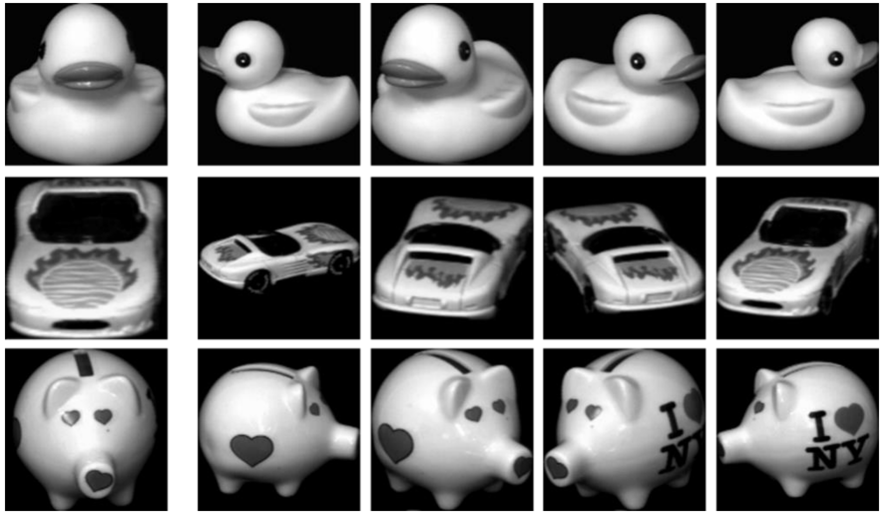
Experiments on Real Surveillance Quality Data

Sample faces of Surveillance Cameras (SC) Face Database [M. Grgic *et al.* 2011].

Method	Rank 1 1 CAM	Rank 1 5 CAM
MDS Learning [S. Biswas *et al.* TPAMI 2013]	30.0	61.1
LSML [M. Kostinger *et al.* CVPR 2012]	64.7	67.2
GMA [A. Sharma *et al.* ICCV 2012]	38.2	50.5
FCPRF + LSML [F. Shen *et al.* PR 2016]	58.0	61.3
SCDL [S. Wang *et al.* CVPR 2012]	48.2	58.5
SCDL + LSML	48.8	60.0
CFDL [D.A. Huang *et al.* ICCV 2013]	45.7	62.2
CFDL + LSML	46.3	63.3
DPFD [ICCV 2015]	69.0	-
Proposed (ADPR)	**73.3**	**-**

Rank-1 accuracy (%) of the proposed approach and comparison with state-of-the-art approaches on the SC face database [M. Grgic *et al.* 2011].

Experiments on Object Recognition Across Pose

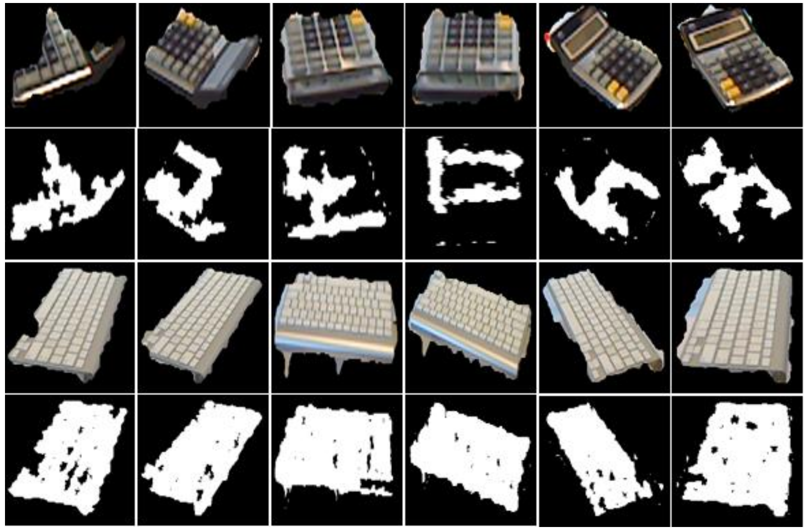


Sample Objects of COIL 20 database [S.A. Nene *et al.* 1996]

Method	Rank-1 Accuracy
MDS Learning [S. Biswas *et al.* TPAMI 2013]	75.6
LSML [M. Kostinger *et al.* CVPR 2012]	80.3
GMA [A. Sharma *et al.* ICCV 2012]	66.1
MvDA [M. Kan *et al* PAMI 2016]	69.7
SCDL [S. Wang *et al.* CVPR 2012]	79.2
SCDL + LSML	82.6
CFDL [D.A. Huang *et al.* ICCV 2013]	78.7
CFDL + LSML	82.0
DPFD [ICCV 2015]	82.2
Proposed (ADPR)	**83.0**

Rank-1 accuracy (%) of the proposed approach and comparison with other approaches on COIL 20 Database [S.A. Nene *et al.* 1996]

Experiments on RGB-D Data



Sample RGB and Depth Objects of RGB-D Database [K. Lai *et al.* 2011].

Method	Visual-Visual	Depth-Depth
MDS Learning [S. Biswas *et al.* TPAMI 2013]	82.2	53.9
LSML [M. Kostinger *et al.* CVPR 2012]	60.1	45.8
GMA [A. Sharma *et al.* ICCV 2012]	70.6	38.9
MvDA [M. Kan *et al* PAMI 2016]	77.2	50.6
SCDL [S. Wang *et al.* CVPR 2012]	80.4	61.1
SCDL + LSML	81.7	62.0
CFDL [D.A. Huang *et al.* ICCV 2013]	81.0	60.5
CFDL + LSML	82.0	61.3
DPFD [ICCV 2015]	86.0	62.0
Proposed (ADPR)	**91.4**	**69.2**

Rank-1 accuracy (%) of the proposed approach and comparison with state-of-the-art approaches on the RGB-D database [K. Lai *et al.* 2011].

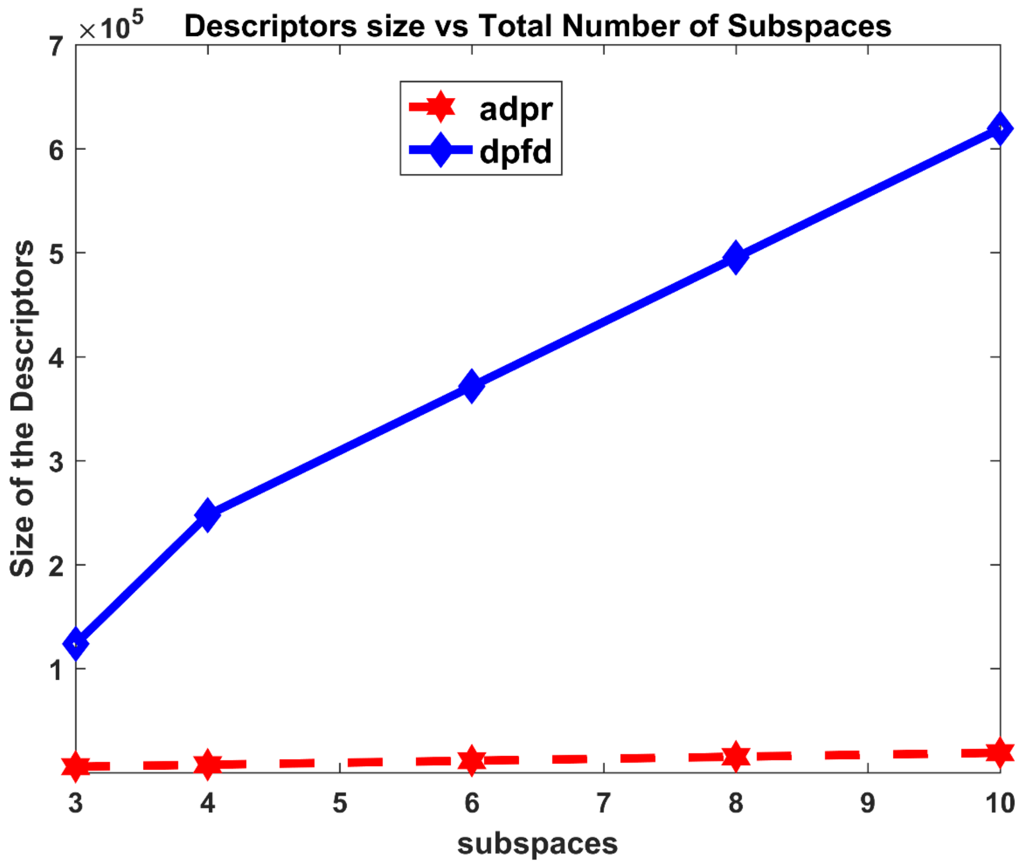
Comparison with deep architectures

Method [Face Recognition]	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

Method [Object Recognition]	Visual-Visual
AlexNet-NN	90.2
AlexNet-ADPR	93.4

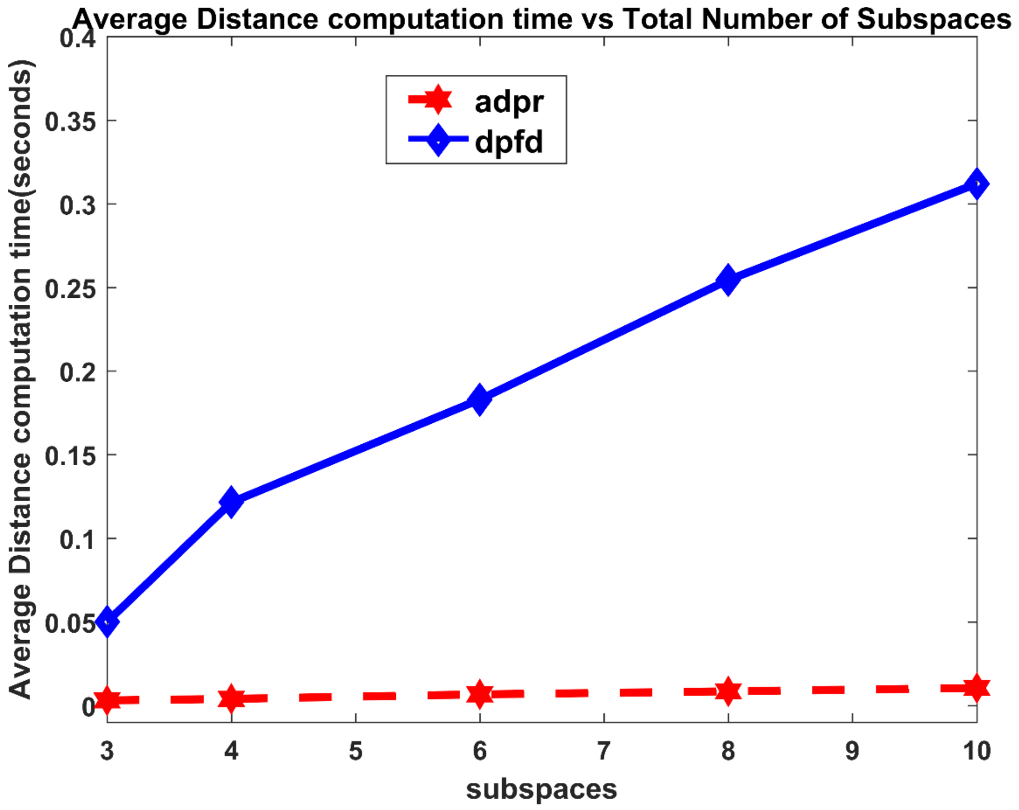
Comparison with VGG [O. M. Parkhi *et al.* 2015] and AlexNet [A. Krizhevsky *et al.* 2012] using their fc6 layer as low level features.
Experiments done on Multi-PIE dataset [R. Gross *et al.* 2007] and RGB-D dataset [K. Lai *et al.* 2011].

Analysis: Descriptor Size vs. Number of Subspaces



Plot of Size of the descriptors vs. Number of Subspaces.
Experiments done on SC Face Database [M. Grgic *et al.* 2011].

Analysis: Distance Computation Time vs. Number of Subspaces



Plot of Distance computation Time vs. Number of Subspaces.
Experiments done on SC Face Database [M. Grgic *et al.* 2011].

Conclusion:

- Novel discriminative pose-free descriptors (DPF) for matching objects across different poses.
- The approaches does not require separate training for different probe poses/viewpoints. This is an advantage over many other approaches which work well when separate training is performed for different poses encountered during testing.
- Very few poses (as little as two/three) are required during the training phase and the method can generalize to unseen poses.

References

- Soubhik Sanyal, Sivaram Prasad Mudunuri, Soma Biswas, Discriminative Pose-Free Descriptors for Face and Object Matching, In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
- Soubhik Sanyal, Devraj Mandal, Soma Biswas, Aligned Discriminative Pose Robust Descriptors for Face and Object Recognition, In Proceedings of the IEEE International Conference in Image Processing (ICIP), 2017.
- Soubhik Sanyal, Sivaram Prasad Mudunuri, Soma Biswas, Discriminative Pose-Free Descriptors for Face and Object Matching, Pattern Recognition, 2017.
- 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.
- M. Kostinger, M. Hirzer, P.Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2288–2295, 2012.
- R. Gopalan, R. Li, R. Chellappa, "Domain Adaptation for Object Recognition: An Unsupervised Approach", IEEE International Conference on Computer Vision, 2011.
- R. Basri, T. Hassner, L. Z. Manor, "A General Framework for Approximate Nearest Subspace Search", IEEE International Conference on Computer Viaion Workshop, 2009
- B. Gong, Y. Shi, F. Sha, K. Grauman, "Geodesic Flow Kernel for Unsupervised Domain Adaptation", IEEE conference on Computer Vision and Pattern Recognition, 2012