Perception Preserving Projections

Saining Xie[1], Jiashi Feng[2], Shuicheng Yan[2], Hongtao Lu[1]

[1] Shanghai Jiao Tong University

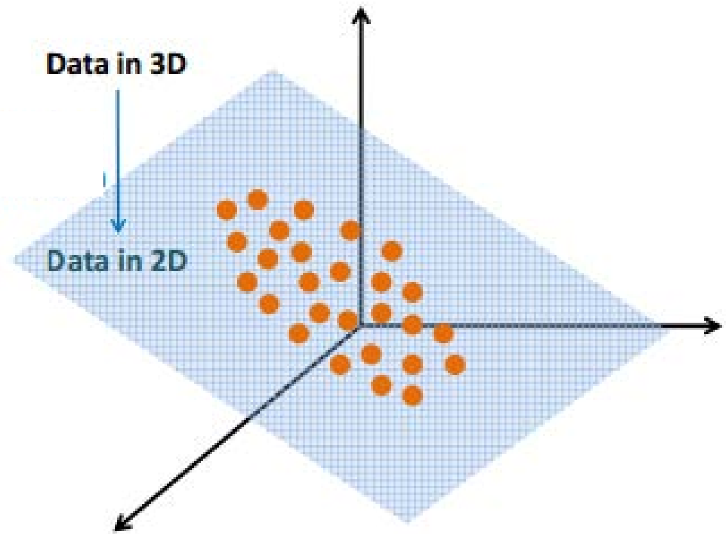
[2] National University of Singapore

Presenter: Chao Zhang

Outline

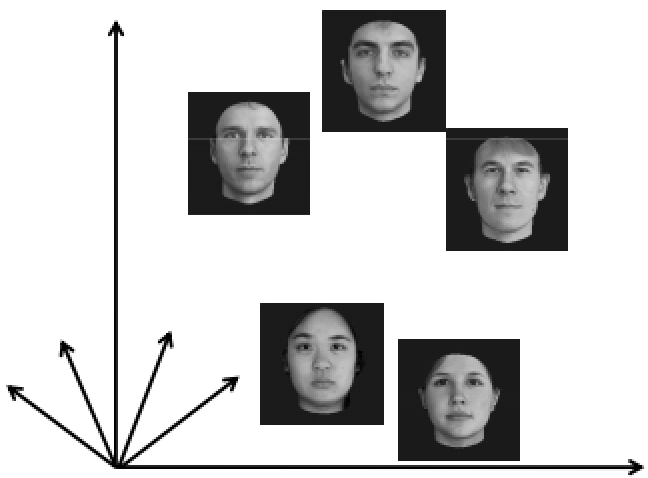
1. Motivations
2. Problem Formulation
3. Optimization
4. Experimental Results
5. Conclusions

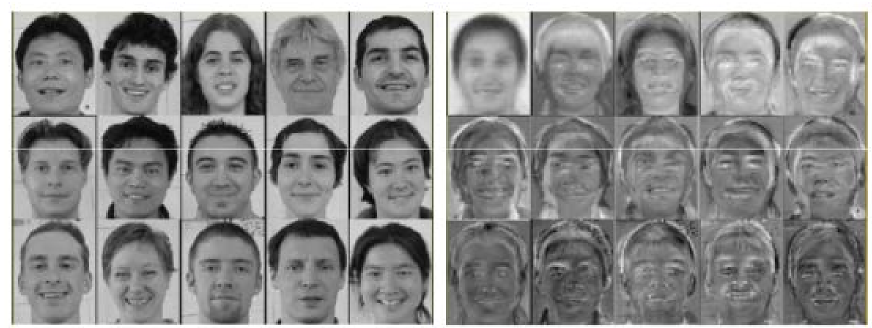
Motivations



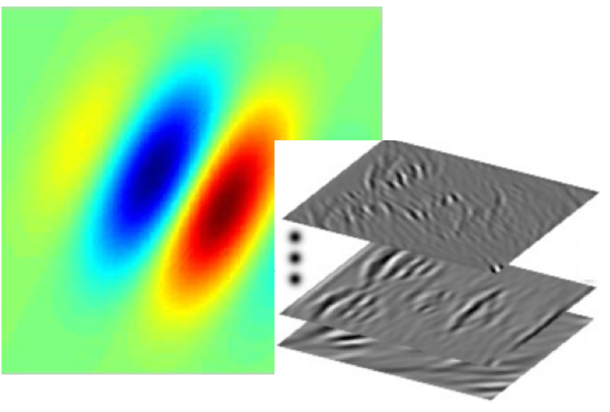
Dimensionality Reduction in Computer Vision

- An image is a point in a high dimensional spac
  - An $n \times m$ image is a point in $\Re^{n \times m}$.

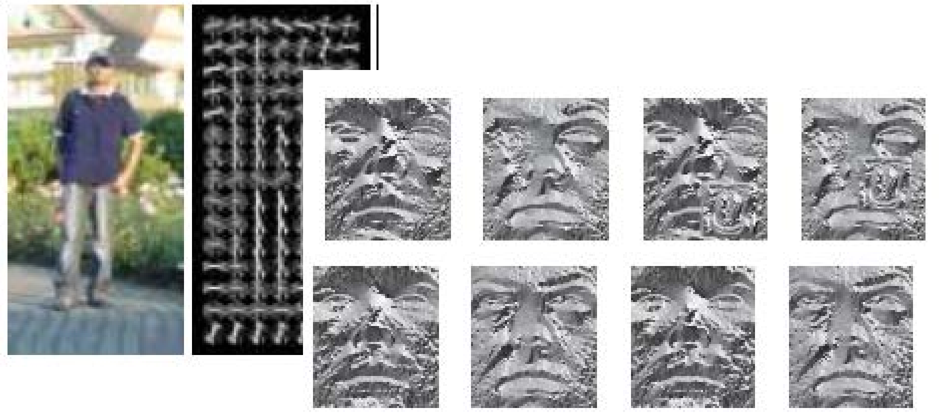




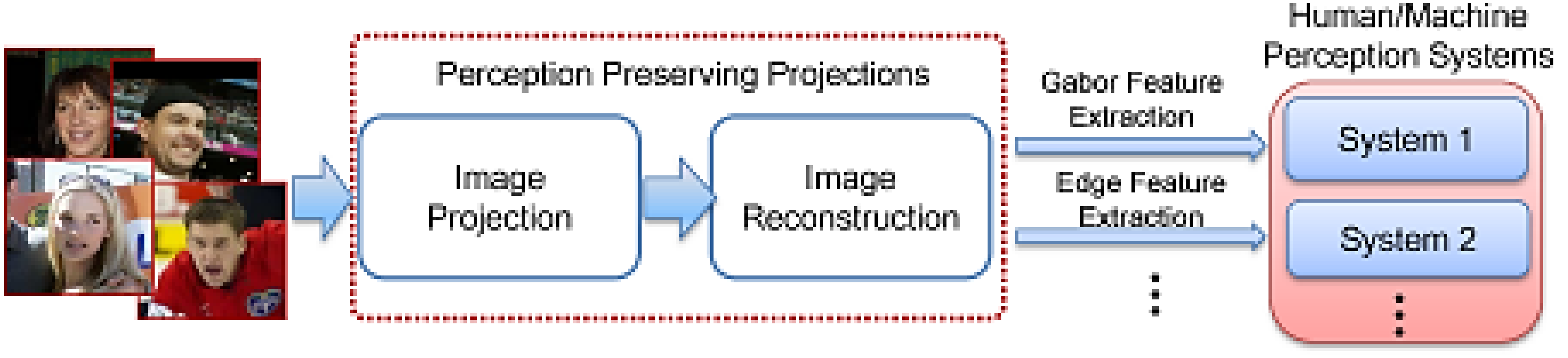
Motivations (cont.)



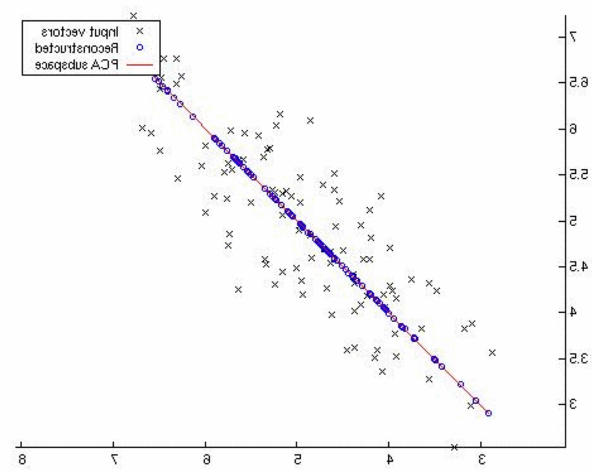
Eg. 2D Gabor Filter



Eg. Gradient Feature



Problem Formulation: *PCA Revisiting*



$$\min_{U^T U = I_r} \mathcal{L}(U) = \|X - UU^T X\|_F^2.$$

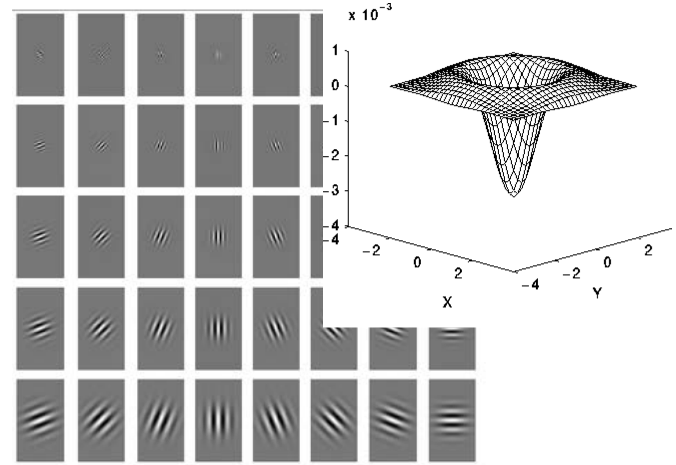
PCA Formulation: Minimize Reconstruction Loss

PCA does NOT take the feature extraction step into consideration, may incur severe information loss in Specific perception systems.

Problem Formulation: *P*erception *P*reserving *P*rojections

Many ***Feature Extractors*** can be expressed as ***Linear Operators***:

1. **Convolution with Linear Filters**
   $f\colon \mathcal{P}(\mathbf{x}) = P_f\mathbf{x} = f * \mathbf{x}.$
2. **Pixel-wise "masking"**
   $P\mathbf{x}$, where P is a $d \times d$ diagonal matrix
3. **Sum of Filters**
   $\sum_{k=1}^{K} P_{f_k}\mathbf{x}$



Objective Function of Perception Preserving Projections:

$$\min_{U^TU=I_r} \mathcal{L}(U) = \|\mathcal{P}'(X) - \mathcal{P}'(UU^TX)\|_F^2,$$

$$P' = [(1-\alpha)P, \alpha I_d]$$

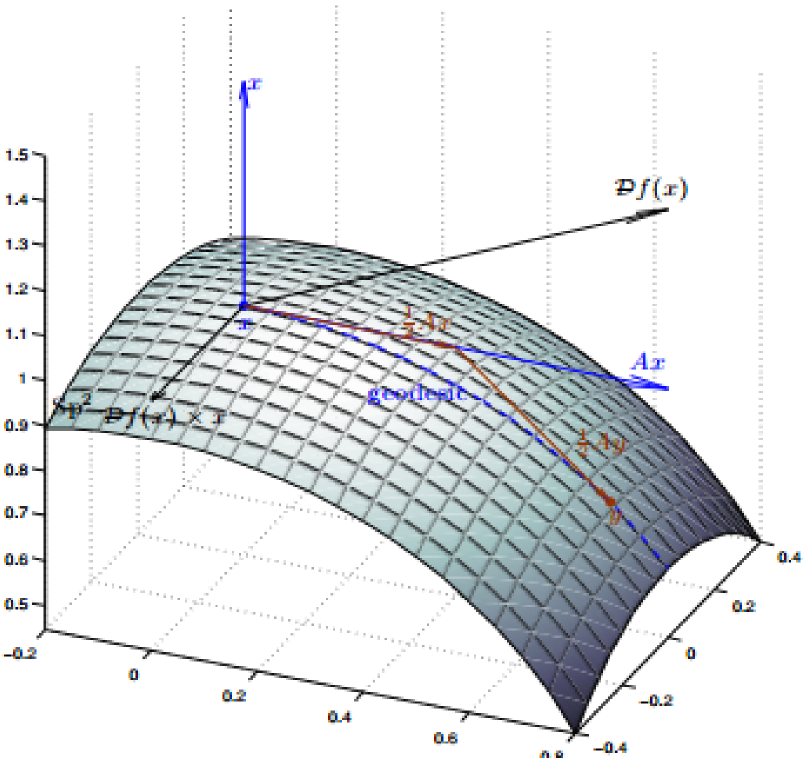
Optimization

Straight-forward solution:
Gradient Descent on Stiefel Manifolds

state-of-the-art off-the-shelf solver:
Cayley Transformation

$$Q = (I - A)(I + A)^{-1}$$

Produce an orthogonal matrix Q.



[2010] **Z. Wen** and W. Yin, A feasible method for optimization with orthogonality constraints, Mathematical Programming

Optimization (cont.)

(Inspired by **Robust PCA** work)

*We can relax the **orthogonal** constraints to a **rank** minimization problem*

$$\min_W \|\mathcal{P}(X) - \mathcal{P}(WX)\|_F^2, \text{ s.t. } \operatorname{rank}(W) \le r,$$

*Since **Nuclear Norm*** is the convex envelope of the rank function.

$$\min_W \|W\|_* + \lambda \|E\|_F^2, \text{ s.t. } \mathcal{P}(X) - \mathcal{P}(WX) = E,$$

Can be solved efficiently by : Alternating Direction Method (ADM)

Problem: $\mathcal{P}(WX) + W = C$ (C is a constant matrix)

Encounter a **Sylvester equation** above in the sub-problem w.r.t variable **W**

Infeasible because O(n^6) time complexity!

*Candès, Emmanuel J., et al. "**Robust principal component analysis?.**" Journal of the ACM (JACM) 58.3 (2011)*

Optimization (cont.)

Solution: Linearize the objective function at the point of W_k

$$\mathcal{L}(W,W_k) = \|W\|_* + \langle Y, -\mathcal{P}(W_kX)\rangle + \mu \left\langle \mathcal{P}^* \left(\mathcal{P}(X) - \mathcal{P}(W^kX) - E\right) X^T, W - W_k \right\rangle + \frac{\mu\eta}{2}\|W - W_k\|_F^2.$$

$$\mathcal{L}(W,Y,W_k) = \|W\|_* + \frac{\mu\eta}{2}\|W - M_k\|_F^2,$$

$$M_k = W_k - \mathcal{P}^*\left(\mathcal{P}(X) - \mathcal{P}(W_kX) - E\right)X^T/\eta + \mathcal{P}^*YX^T/\mu\eta$$

CLOSED FORM solution (Soft-thresholding):

$$W_{k+1} = U\mathcal{S}_{\frac{1}{\mu\eta}}(\Sigma)V^T, \qquad \mathcal{S}_\varepsilon[x] = sgn(x)max(|x| - \varepsilon, 0)$$

Speed-up Tricks:

1. adaptive penalty strategy.

$$\mu_{k+1} = \begin{cases} \rho_0\mu_k, & \text{if } \mu_k \max(\sqrt{\eta}\varepsilon_W, \varepsilon_E)/\|\mathcal{P}(X)\| < \varepsilon_2, \\ \mu_k, & \text{otherwise.} \end{cases}$$

2. PROPACK for partial SVD

3. Skinny SVD to avoid full matrix multiplication

Lin, Z.,et al. Linearized alternating direction method with adaptive penalty for low-rank representation, NIPS 2011

Experimental Results: Synthetic Data

randomly generate
100 gray level patches
with Vertical lines and Horizontal outliers

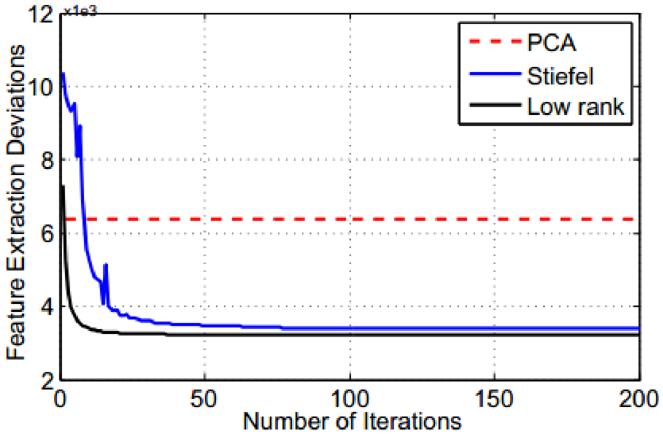
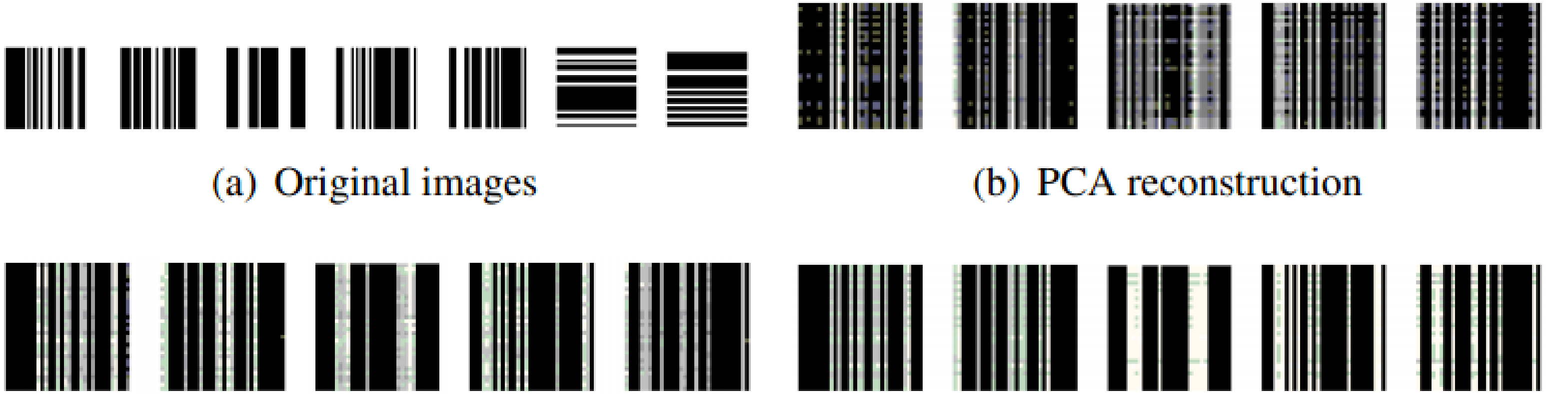


Figure 2: Feature deviation descent curve on the synthetic data along with iterations.



(a) Original images

(b) PCA reconstruction

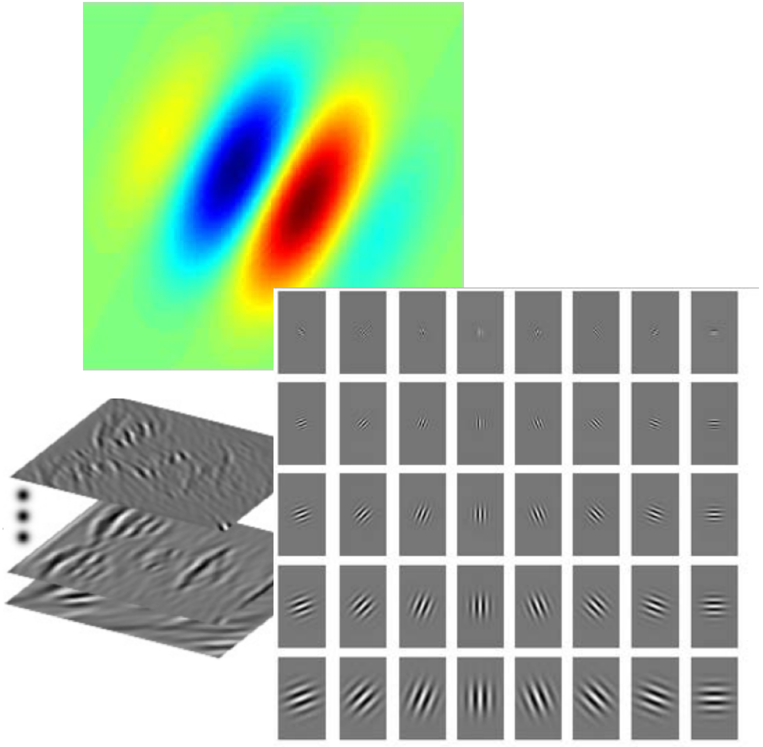
(c) PPP reconstruction based on gradient descent over Stiefel manifold

(d) PPP reconstruction based on low rank optimization

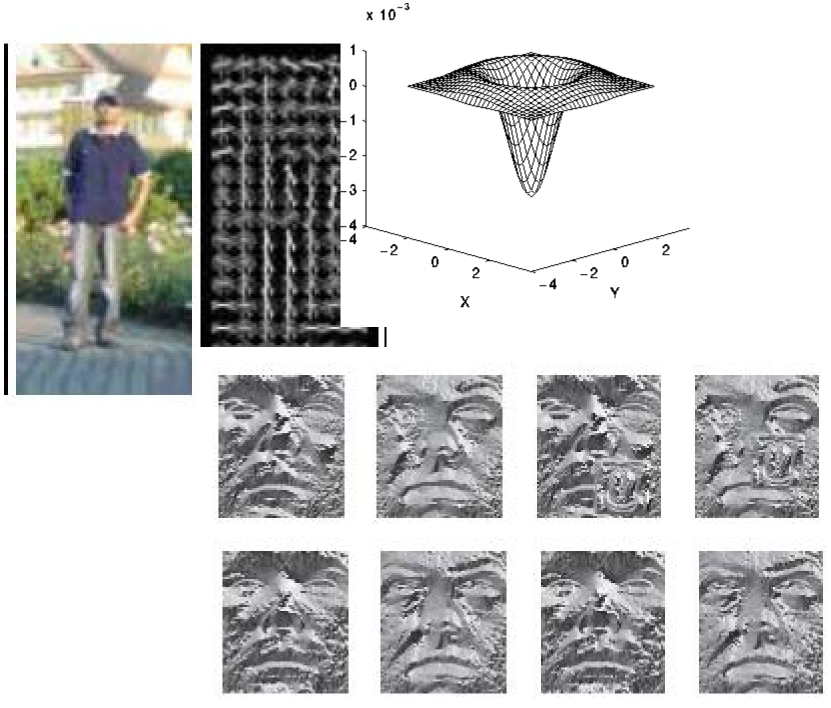
Figure 4: Examples of the synthetic data and reconstruction results from different methods.

Experimental Results:

Gradient Preserving and Gabor Feature Preserving



Eg. 2D Gabor Filter

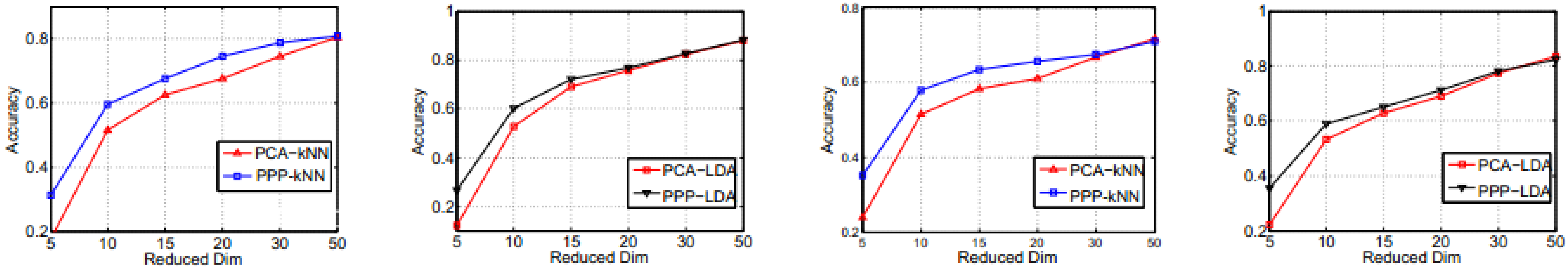


Eg. Gradient Feature

Experimental Results:

Gradient and Gabor Feature Preserving (cont.)

Figure: Results on FRGC dataset (with multiple classifiers)



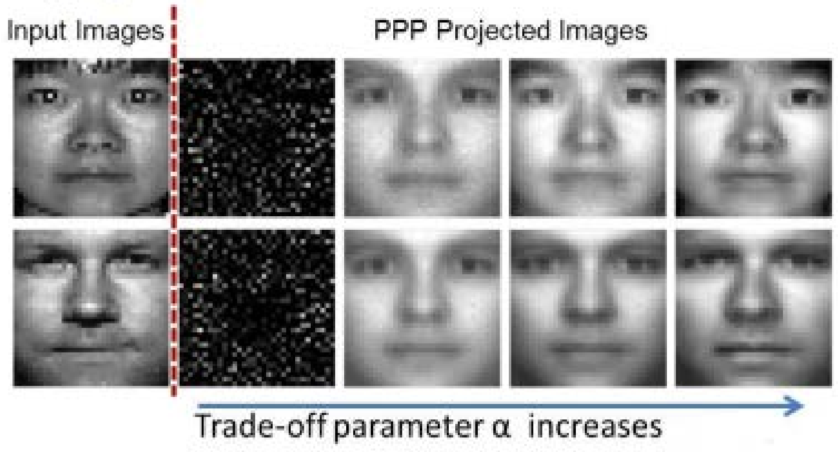
(a) Gabor + kNN (b) Gabor + LDA (c) LoG + kNN (d) LoG + LDA

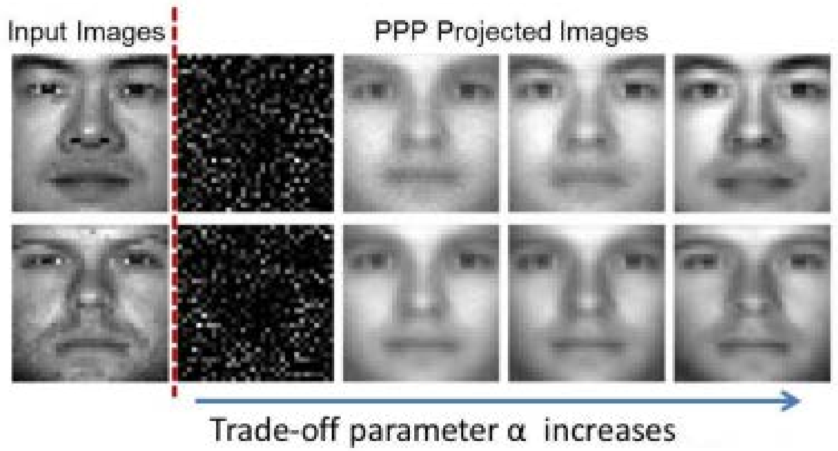
Table: Results on Extended Yale-B: Gabor Feature (above) / Gradient Feature (bottom)

Dim	PCA+kNN	PPP-S+kNN	PPP-L+kNN	PCA+LDA	PPP-S+LDA	PPP-L+LDA
5	5.82 ± 1.64	**43.51 ± 2.63**	42.30 ± 2.17	10.40 ± 1.53	**44.30 ± 1.17**	44.41 ± 2.68
10	33.56 ± 1.24	68.34 ± 2.81	**70.58 ± 2.29**	49.89 ± 2.17	69.57 ± 2.60	**71.14 ± 2.73**
15	48.66 ± 1.95	**77.63 ± 1.25**	76.52 ± 1.09	70.13 ± 1.07	**78.75 ± 2.73**	77.40 ± 1.66
20	58.39 ± 2.49	**81.66 ± 2.83**	80.21 ± 2.10	80.54 ± 2.79	83.00 ± 1.69	**83.54 ± 2.96**
30	71.36 ± 2.07	84.23 ± 2.26	**84.34 ± 3.03**	85.12 ± 1.35	85.12 ± 2.85	**86.92 ± 1.09**
50	79.08 ± 2.27	82.66 ± 1.20	**86.13 ± 0.81**	89.15 ± 2.44	86.02 ± 1.75	**89.58 ± 1.16**

Dim	PCA+kNN	PPP-S+kNN	PPP-L+kNN	PCA+LDA	PPP-S+LDA	PPP-L+LDA
5	4.47 ± 2.92	40.04 ± 3.38	**40.16 ± 0.69**	17.34 ± 1.93	36.13 ± 3.17	**38.59 ± 1.77**
10	41.05 ± 2.93	64.21 ± 1.87	**69.46 ± 1.64**	62.53 ± 2.04	**74.83 ± 3.28**	72.71 ± 2.62
15	57.94 ± 1.32	75.50 ± 4.40	**79.08 ± 2.79**	82.21 ± 2.22	84.56 ± 1.15	**87.81 ± 1.68**
20	70.36 ± 2.94	78.97 ± 0.74	**81.77 ± 1.25**	87.92 ± 1.04	86.80 ± 2.21	**91.05 ± 1.38**
30	79.19 ± 2.91	79.42 ± 2.36	**82.89 ± 1.86**	90.72 ± 1.28	90.04 ± 2.06	**91.39 ± 1.77**
50	83.00 ± 1.97	82.44 ± 2.13	**83.11 ± 0.95**	91.50 ± 2.26	92.06 ± 2.76	**92.17 ± 2.45**

Experimental Results:
Gradient and Gabor Feature Preserving (cont.)





when alpha is small, the reconstructed images of different persons look quite similar. However, by extracting the Gabor features, these images can be distinguished correctly.

Conclusion

- Proposed the **perception preserving projection** method, which is able to **preserve** the important information for specific perception system in the image projection process.
- explicitly **embed** the feature preserving metric provided by a certain type of perception systems into the loss function.
- The results suggest PPP can better preserve the **discriminative** and **domain-specific** feature information.

☹ the current framework is quite *naïve.*

Future work:

•How can PPP be regarded as an implementation of ***unsupervised joint embedding*** of different domains?

•Extend the range of its application scenarios.

Thank You

Q&A: xiesaining@gmail.com