Metric learning method for subspace-based image-set recognition

Naoya Sogi, Lincon Souza, Bernardo Gatto, Kazuhiro Fukui

Graduate School of Systems and Information Engineering, University of Tsukuba, Japan

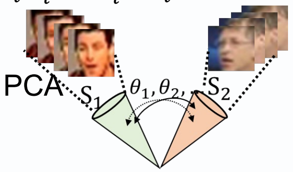
Center for Artificial Intelligence Research, University of Tsukuba, Japan

(1) Introduction

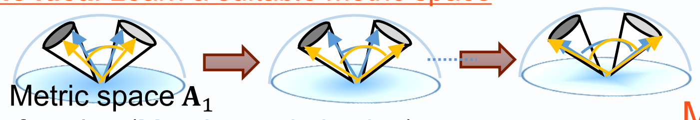
- *Image-set recognition*:
  - ➢ Efficient approach for object recognition.
  - ➢ Utilizing multiple images (multi-view, video) as an input.
- *Subspace-method*:
  - ➢ General approach for the image-set recognition.
  1. Representing an image-set by a low-dimensional subspace.
  2. Classifying an input image-set by the subspace-similarity
- ***Our objective***: improving performance of subspace-based methods
- ***Approach***: constructing a metric learning method with subspaces

(2) Proposed method

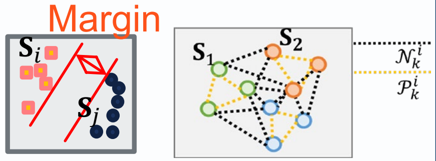
- Subspace similarity with a metric
  - $\mathrm{sim}(\mathbf{S}_1,\mathbf{S}_2) = \|\hat{\mathbf{S}}_1^T\mathbf{A}\hat{\mathbf{S}}_2\|_F^2 = \sum_j \cos^2\theta_j$, $\hat{\mathbf{S}}_i = \mathbf{S}_i\mathbf{U}_i\Sigma_i^{-1/2}$, $\mathbf{U}_i\Sigma_i\mathbf{U}_i^T = \mathbf{S}_i^T\mathbf{A}\mathbf{S}_i$.
  - θ_j: j-th canonical angle
  - $\mathbf{S}_i$: Basis of a subspace generated by PCA



- **Basic idea: Learn a suitable metric space**



- Cost function (Margin maximization)

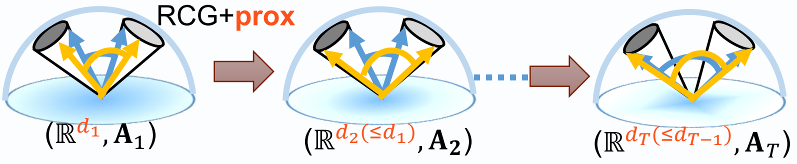


- $\mathbf{A}^* = \min_{\mathbf{A}} J(\mathbf{A}) = J_b(\mathbf{A}) + J_w(\mathbf{A}) + \lambda\left(1 - \|\mathbf{A}\|_F^2\right)^2$,
- $J_b(\mathbf{A}) = \frac{1}{kN}\sum_{i=1}^N \sum_{j\in\mathcal{N}_k^i} \mathrm{sim}(\mathbf{S}_i,\mathbf{S}_j)$,
- $J_w(\mathbf{A}) = \frac{1}{kN}\sum_{i=1}^N \sum_{j\in\mathcal{P}_k^i} d^2(\mathbf{S}_i,\mathbf{S}_j) = \frac{1}{kN}\sum_{i=1}^N \sum_{j\in\mathcal{P}_k^i}(r - \mathrm{sim}(\mathbf{S}_i,\mathbf{S}_j))$
- $\mathcal{N}_k^i$/ $\mathcal{P}_k^i$:k-neighborhood of the i-th subspace
- Optimization:
  - Use Riemannian conjugate gradient, RCG. ($\mathbf{A}$ must be a SPD matrix)
  - Need a gradient of sim().
    - Reformulate

$$\begin{aligned}\mathrm{sim}(\mathbf{S}_i,\mathbf{S}_j) &= \|\hat{\mathbf{S}}_i^T\mathbf{A}\hat{\mathbf{S}}_j\|_F^2 = tr(\hat{\mathbf{S}}_j^T\mathbf{A}\hat{\mathbf{S}}_i\hat{\mathbf{S}}_i^T\mathbf{A}\hat{\mathbf{S}}_j) = tr(\hat{\mathbf{S}}_j\hat{\mathbf{S}}_j^T\mathbf{A}\hat{\mathbf{S}}_i\hat{\mathbf{S}}_i^T\mathbf{A})\\ &= tr(\mathbf{S}_j\mathbf{U}_j\Sigma_j^{-1/2}\Sigma_j^{-1/2}\mathbf{U}_j^T\mathbf{S}_j^T\mathbf{A}\mathbf{S}_i\mathbf{U}_i\Sigma_i^{-1/2}\Sigma_i^{-1/2}\mathbf{U}_i^T\mathbf{S}_i^T\mathbf{A})\\ &= tr(\mathbf{S}_i^T\mathbf{A}\mathbf{S}_j(\mathbf{S}_j^T\mathbf{A}\mathbf{S}_j)^{-1}\mathbf{S}_j^T\mathbf{A}\mathbf{S}_i(\mathbf{S}_i^T\mathbf{A}\mathbf{S}_i)^{-1}).\end{aligned}$$

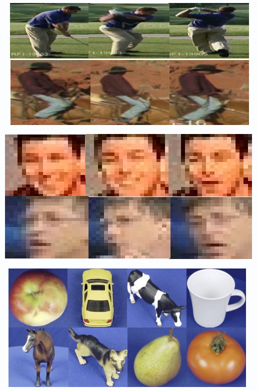
(3) Extension method

- Rank($\mathbf{A}$)=dim. of the metric space→can estimate the optimal dim automatically by low-rank constraint
- Cost function: $\min_{\mathbf{A}} J(\mathbf{A}) + \eta\|\mathbf{A}\|_*$, $\|\ \|_*$trace norm.
- Optimization: RCG + proximal gradient method
  - The trace norm constraint is not differentiable.
  - Apply the following operation after RCG step
  - $\mathrm{prox}_\eta^{tr}(\mathbf{A}) = \mathbf{U}\max(\mathbf{\Sigma} - \eta\mathbf{I},\mathbf{0})\mathbf{U}^{\mathrm{T}}, \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}} = \mathbf{A}$



(4) Results

- Datasets:1. UCF sport(10 actions) 2. YTC(47 faces) 3.ETH(8 objects)



model	method	UCF	YTC	ETH
Affine hull	AHISD[1]	53.33	56.31±4.61	70.25±4.78
subspace	GGDA[2]	47.33	47.45±8.06	93.75±4.75
	PML[3]	72.67	54.82±4.63	93.75±3.17
	RMML[4]	59.33	54.96±4.96	90.25±4.16
	Proposed	71.33	59.29±2.72	**95.28±3.21**
	+$\|\mathbf{A}\|_*$	**74.00**	**59.85±3.53**	95.25±2.82

(5) Conclusion

- We introduced a metric space into a subspace-method and constructed metric learning frameworks by efficiently handing subspaces.
- We showed that our methods can achieve competitive results compared with various previous methods.

[1] Cevikalp, H. and Triggs, B. "Face recognition based on image sets", CVPR 2010.

[2] Harandi, M.T. et al, "Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching", CVPR 2011.

[3] Huang, Z. et al, "Projection metric learning on Grassmann manifold with application to video based face recognition", CVPR 2015.

[4] Zhu, P. et al, "Towards Generalized and Efficient Metric Learning on Riemannian Manifold. IJCAI 2018.