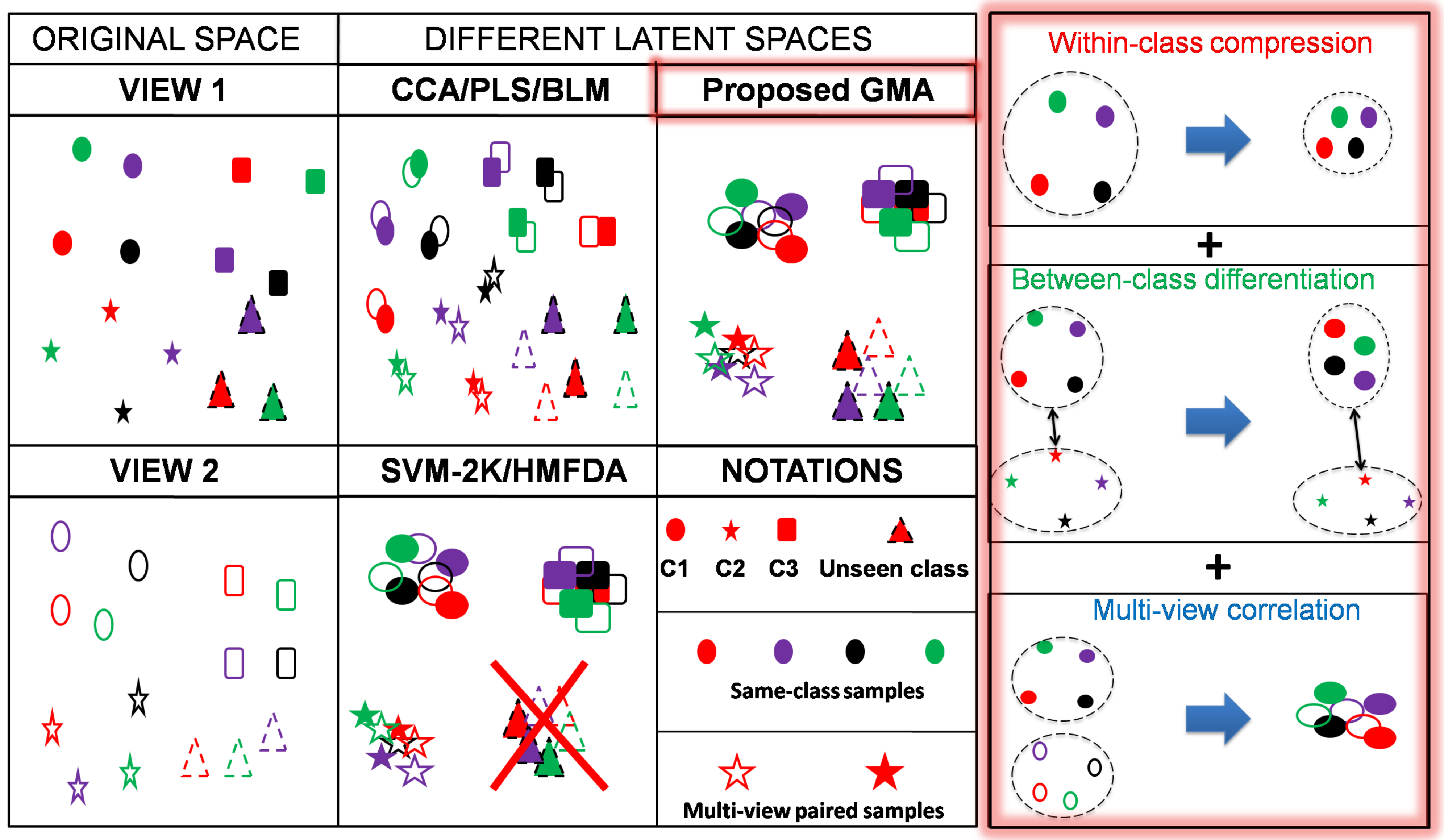
Generalized Multiview Analysis: A Discriminative Latent Space

Abhishek Sharma Abhishek Kumar Hal Daumé III David W. Jacobs
Computer Science Department, University of Maryland, College Park

Contribution

We introduce a multi-view, supervised, domain independent, generalizable to unseen classes and kernelizable mapping to latent space to facilitate multi-view classification and retrieval. Our method has a close-form solution. We show significant improvements over CCA and other domain dependent approaches.

Compression + Discrimination + Correlation = GMA



Mathematical Formulation

CCA: $argmax_{v_1,v_2}\ v_1^T X_1 X_2^T v_2$ s.t. $v_1^T X_1 X_1^T v_1 = v_2^T X_2 X_2^T v_2 = 1$

CCA+Compression: $argmax_{v_1,v_2}\ v_1^T X_1 X_2^T v_2$ s.t. $v_1^T B_1 v_1 = v_2^T B_2 v_2 = 1$

B_i is Within-class variance: LDA, MFA, LPP and so on

Convex relaxation: $argmax_{v_1,v_2}\ v_1^T X_1 X_2^T v_2$ s.t. $v_1^T B_1 v_1 + \alpha v_2^T B_2 v_2 = 1$

GMA: $argmax_{v_1,v_2}\ v_1^T A_1 v_1 + \mu v_2^T A_2 v_2 + \beta v_1^T Z_1 Z_2^T v_2$ s.t. $v_1^T B_1 v_1 + \alpha v_2^T B_2 v_2 = 1$

Z_i are exemplars in view i and A_i is Between-class variance: LDA, MFA and so on

Make the Lagrangian > differentiate and equate to 0 > solve eigen-value problem :)

Closed form solution: $$\begin{bmatrix} A_1 & \beta Z_1 Z_2^T \\ \beta Z_2 Z_1^T & \mu A_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} B_1 & 0 \\ 0 & \alpha B_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}$$

Unified view and supervised extensions under GMA

GMPCA: $A_i = X_i X_i^T/N_i, B_i = I, Z_i = X_i$; **BLM**: $A_i = X_i X_i^T/N_i, B_i = I, Z_i = X_i$

CCA: $A_i = 0, B_i = X_i X_i^T, Z_i = X_i$; **PLS**: $A_i = 0, B_i = I, Z_i = X_i$

GMLDA: $A_i = S_i^B, B_i = S_i^W, Z_i = M_i\ or\ X_i$, where S^W/S^B are Within/Between -class scatter matrices and M_i is the class mean matrix for view i

Mathematical Formulation continued..

GMMFA: $A_i = S_{(B)i}, B_i = S_{(W)i}, Z_i = X_i$, where $S^{kk}_{(B/W)i} = \sum_{kl,k\neq l} W^{kl}_{(b/w)i}$. The *within class compression* or *intrinsic graph*- $W^{kl}_{wi} = \begin{cases} 1 & : k \in R_i^{k1}(l)\ or\ l \in R_i^{k1}(k) \\ 0 & : \text{otherwise} \end{cases}$, where $R_i^{k1}(l)$ indicates the index set of the $k1$ nearest neighbors of the sample x_i^l in the same class. The *between class separation* or *penalty* graph- $W^{kl}_{bi} = \begin{cases} 1 & : (k,l) \in P_i^{k2}(c_l)\ or\ (k,l) \in P_i^{k2}(c_k) \\ 0 & : \text{otherwise} \end{cases}$, where $P_i^{k2}(l)$ is a set of data pairs that are the $k2$ nearest pairs among the set $\{(k,l) : k \text{ and } l \text{ are not in the same class}\}$.

Algorithmic view of the overall framework

Step 1: Input multi-view paired data X_i with label/similarity information.
Step 2: Learn projection directions v_i using GMA.
Step 3: Project X_i to latent space using v_i.
Step 4: Use k-NN matching or learn a classifier in the latent space.

Experimental Results

MultiPIE pose and lighting invariant face recognition

Manually annotated fiducial points for cropping are publicly available[1], simple intensity used as feature.
Mode 1- Training: 129 subjects, 5 lighting; Testing: Same 129 subjects, 18 lightings, frontal pose gallery in illum 7.
Mode 2- Training: 129 subjects, 5 lighting; Testing: Different 120 subjects, 18 lightings, frontal pose gallery in illum 7.

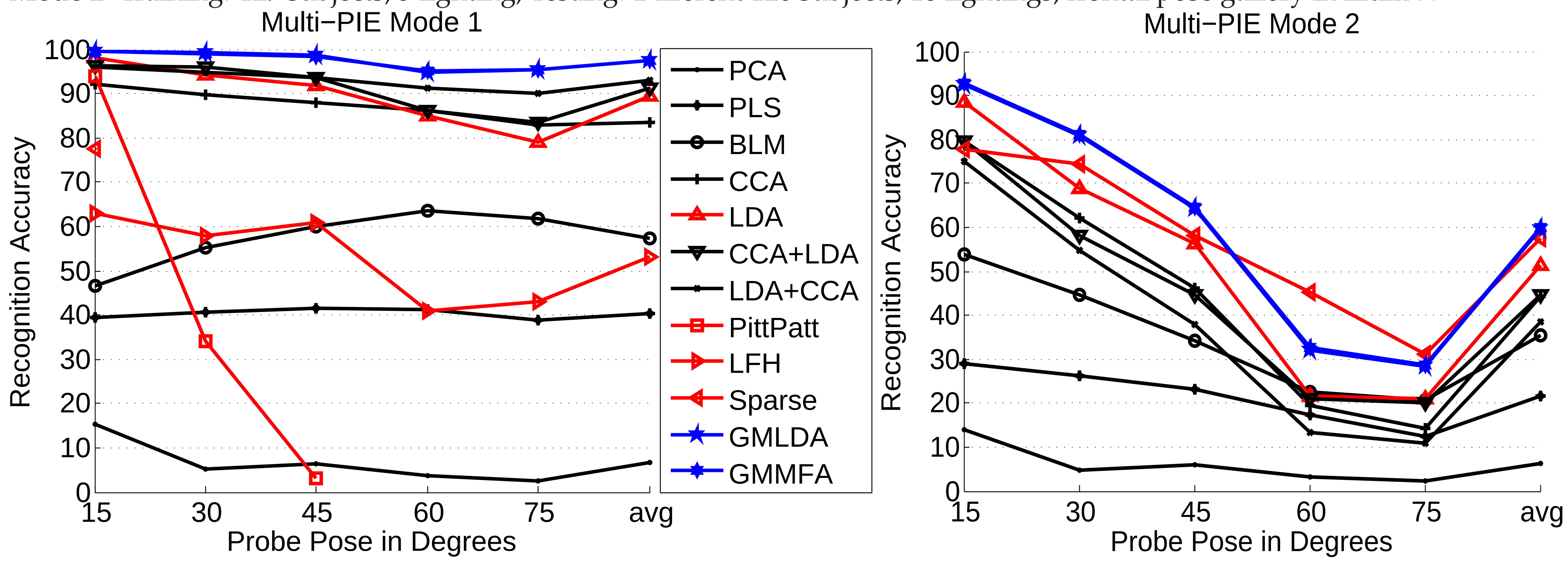


Fig 1 Recognition accuracy for MultiPIE. Red curves indicate domain dependent methods under multi-view setting.

Text-Image retrieval

Wikipedia text-image data: 10 classes, 2173/693 training/testing, publicly available features[2]
Pascal image-tag data: 20 classes, 2808/2841 training/testing, publicly available features[3]

Query	Wikipedia Dataset							Pascal Dataset			
	Others					Proposed		Otheres		Proposed	
	PLS	BLM	CCA	SM	SCM	GMMFA	GMLDA	KPLS	KCCA	KGMLDA	KGMMFA
Image	0.207	0.237	0.182	0.225	**0.277**	0.264	0.272	0.279	0.298	0.421	**0.427**
Text	0.192	0.144	0.209	0.223	0.226	0.231	**0.232**	0.232	0.269	0.328	0.339
Average	0.199	0.191	0.196	0.224	**0.252**	0.248	**0.253**	0.256	0.283	0.375	**0.383**

[1] http://www.umiacs.umd.edu/~bhokaal/data/FERET_MultiPIE_fiducials.tar
[2] http://www.svcl.ucsd.edu/projects/crossmodal/ [3] http://www.cs.utexas.edu/~grauman/research/datasets.html