Visual Feature Extraction by Unified Discriminative Subspace Learning

Yun (Raymond) Fu

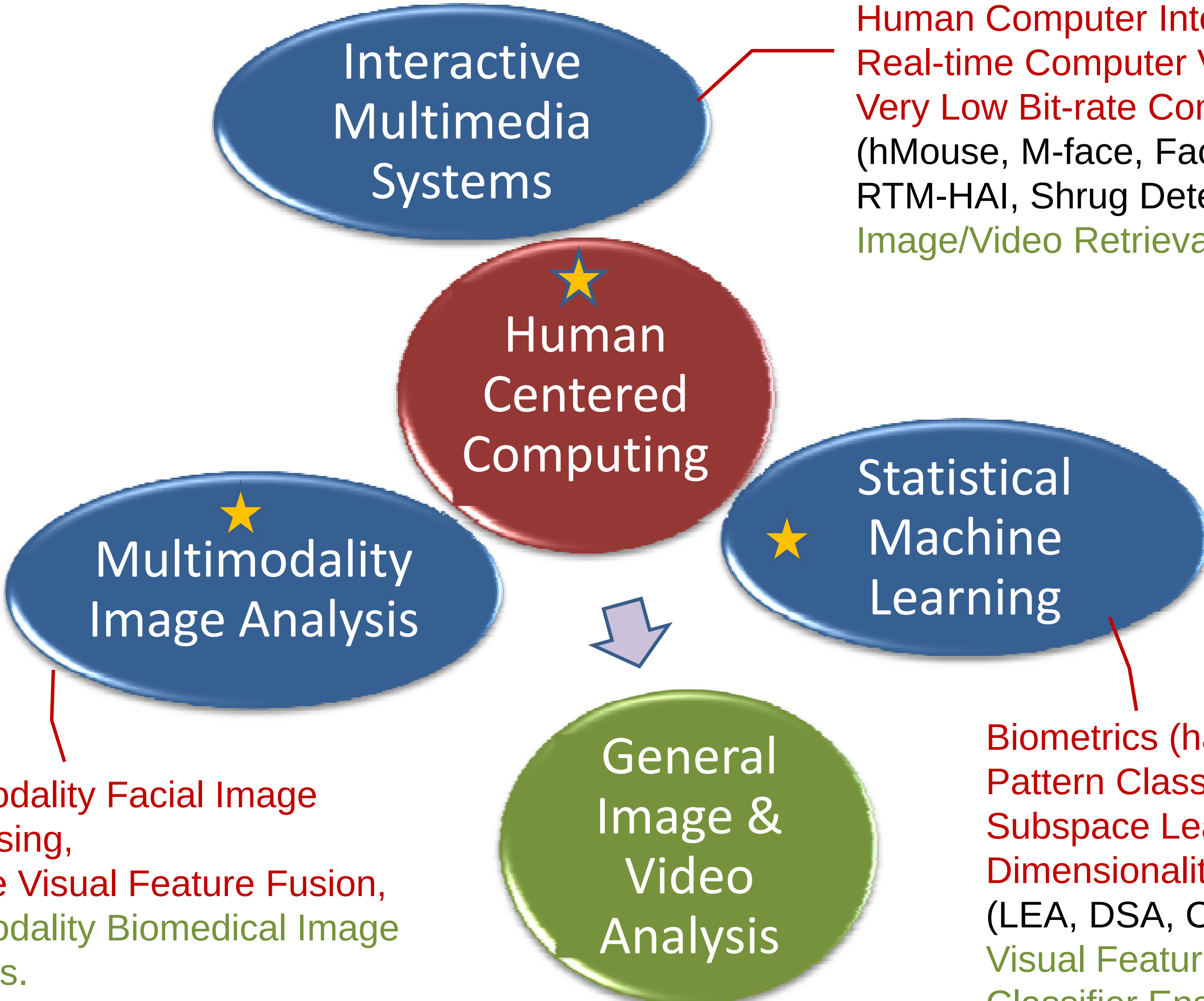
Dr. and Scientist, BBN Technologies
Lecturer, Computer Science, Tufts University

University of Illinois at Urbana-Champaign

My Background

- Ph.D, ECE, UIUC, 2004-2008
  Beckman Graduate Fellow, BI, 2007-2008
  Research Assistant, ECE & BI & CSL, 2004-2007
- M.S., Statistics, UIUC, 2007
- M.S., Pattern Recognition & Intelligent Systems, School of Electronics and Information Engineering, Xi'an Jiaotong University, China, 2004
  Research Assistant, Institute of Artificial Intelligence & Robotics, 2001-2004
- B.E. ,with highest honor, Information Engineering, XJTU, 2001

Research Interests

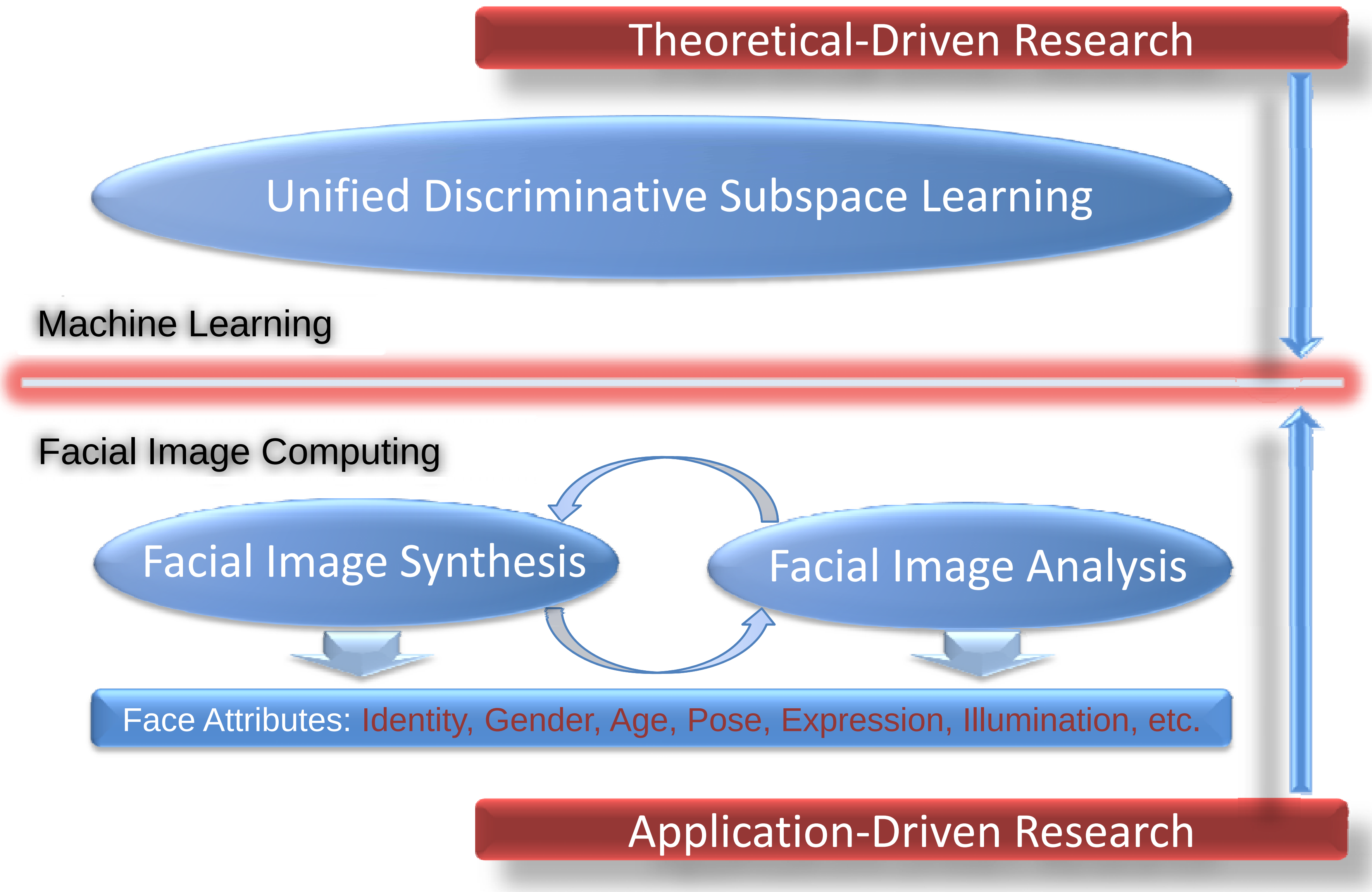


Human Computer Interaction,
Real-time Computer Vision systems,
Very Low Bit-rate Communication.
(hMouse, M-face, Facetransfer, RTM-HAI, Shrug Detector, EAVA, etc.)
Image/Video Retrieval/Indexing.

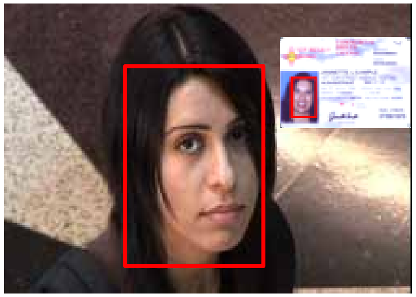
Multimodality Facial Image Processing,
Multiple Visual Feature Fusion,
Multimodality Biomedical Image Analysis.

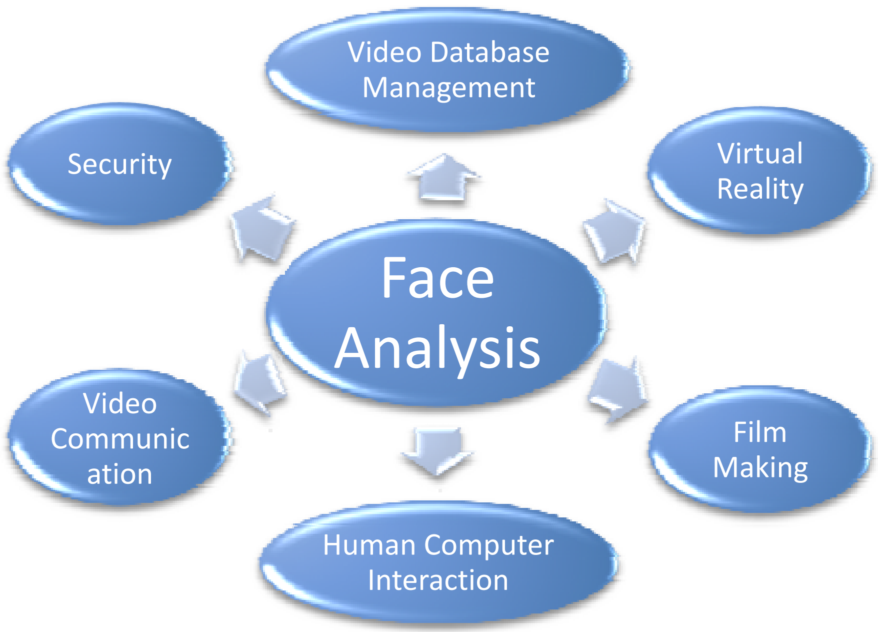
Biometrics (hard/soft),
Pattern Classification,
Subspace Learning,
Dimensionality Reduction.
(LEA, DSA, CEA, CTA, etc.)
Visual Feature Extraction,
Classifier Ensemble.

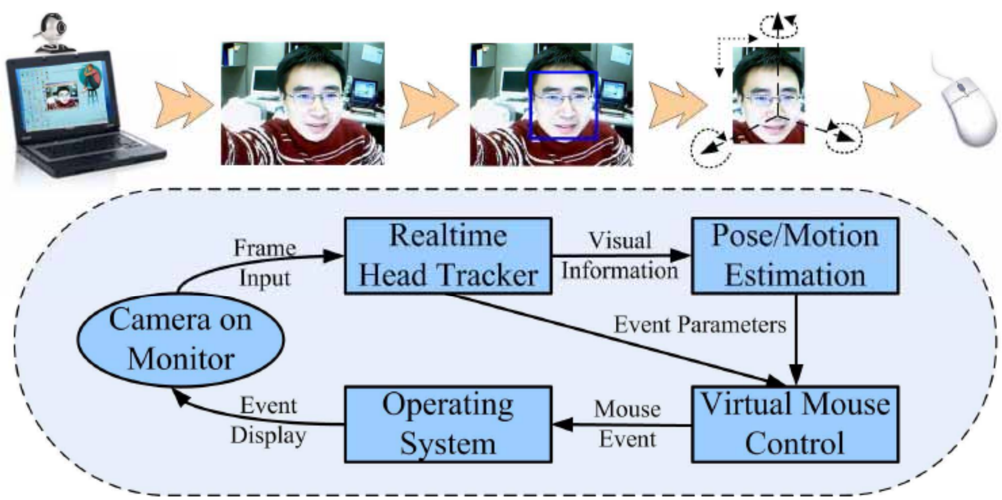
Research Overview and Motivation



The Importance of Faces

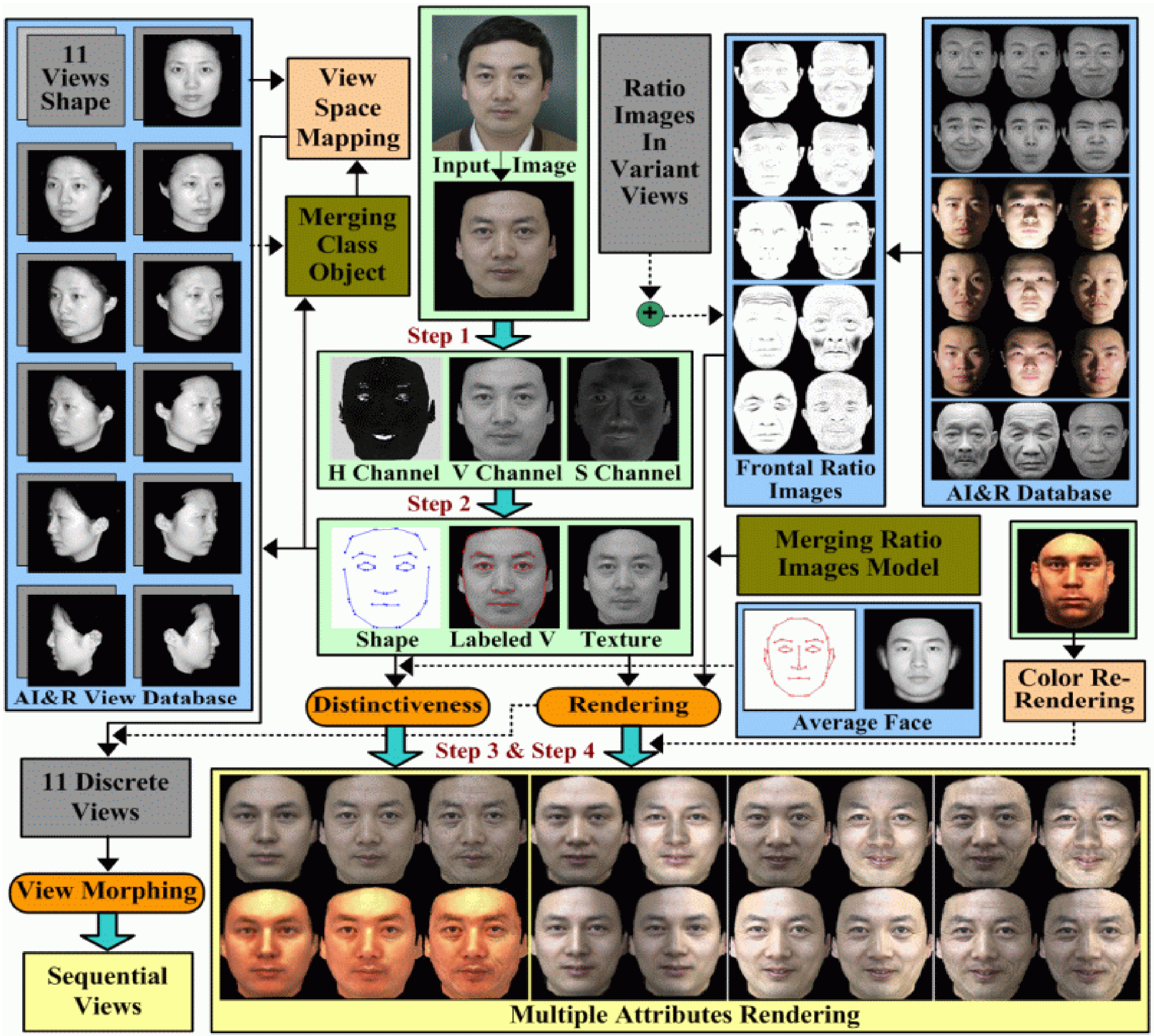




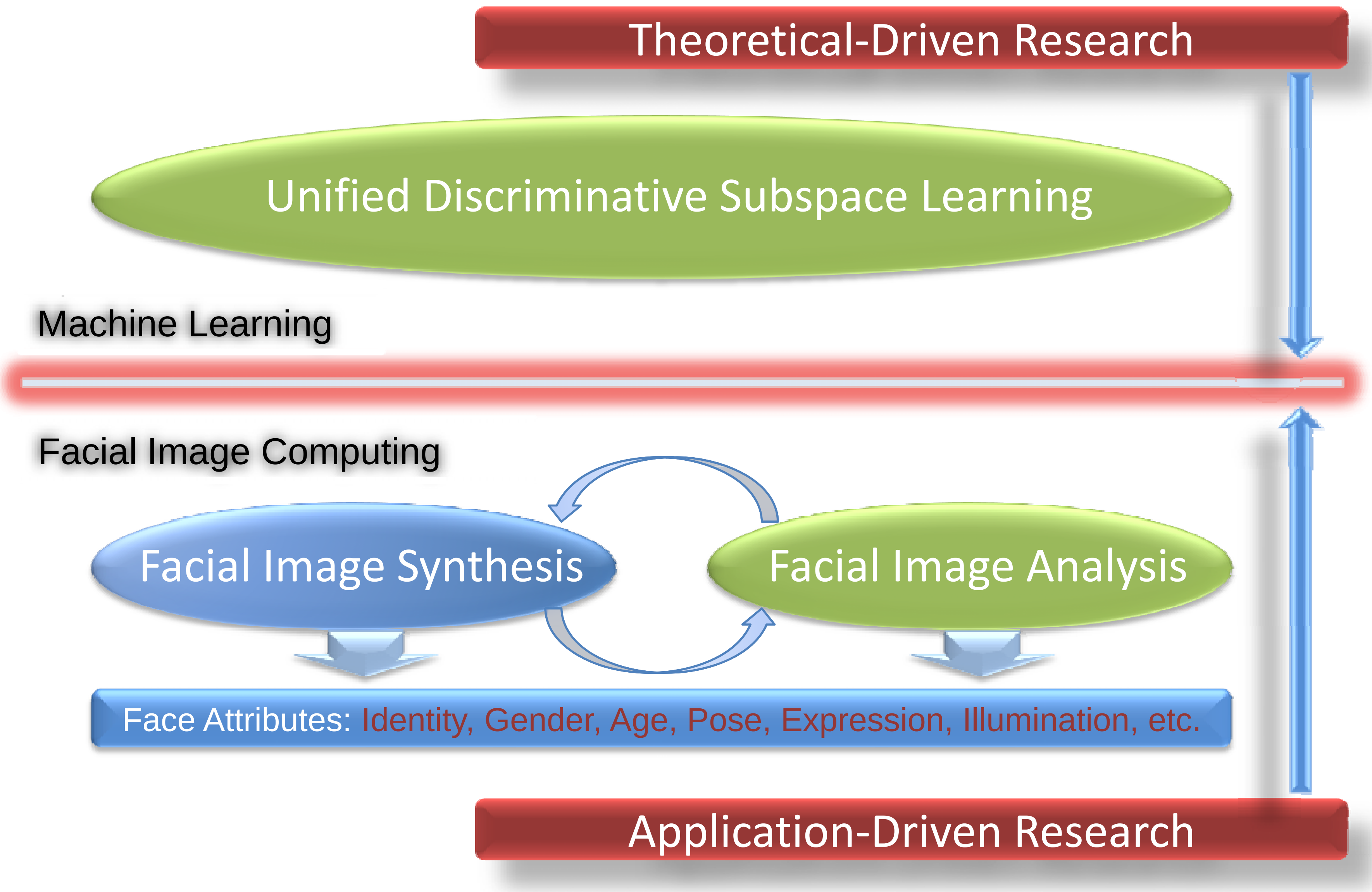


Courtesy of Jilin Tu

Framework of Facial Image Synthesis



Research Overview



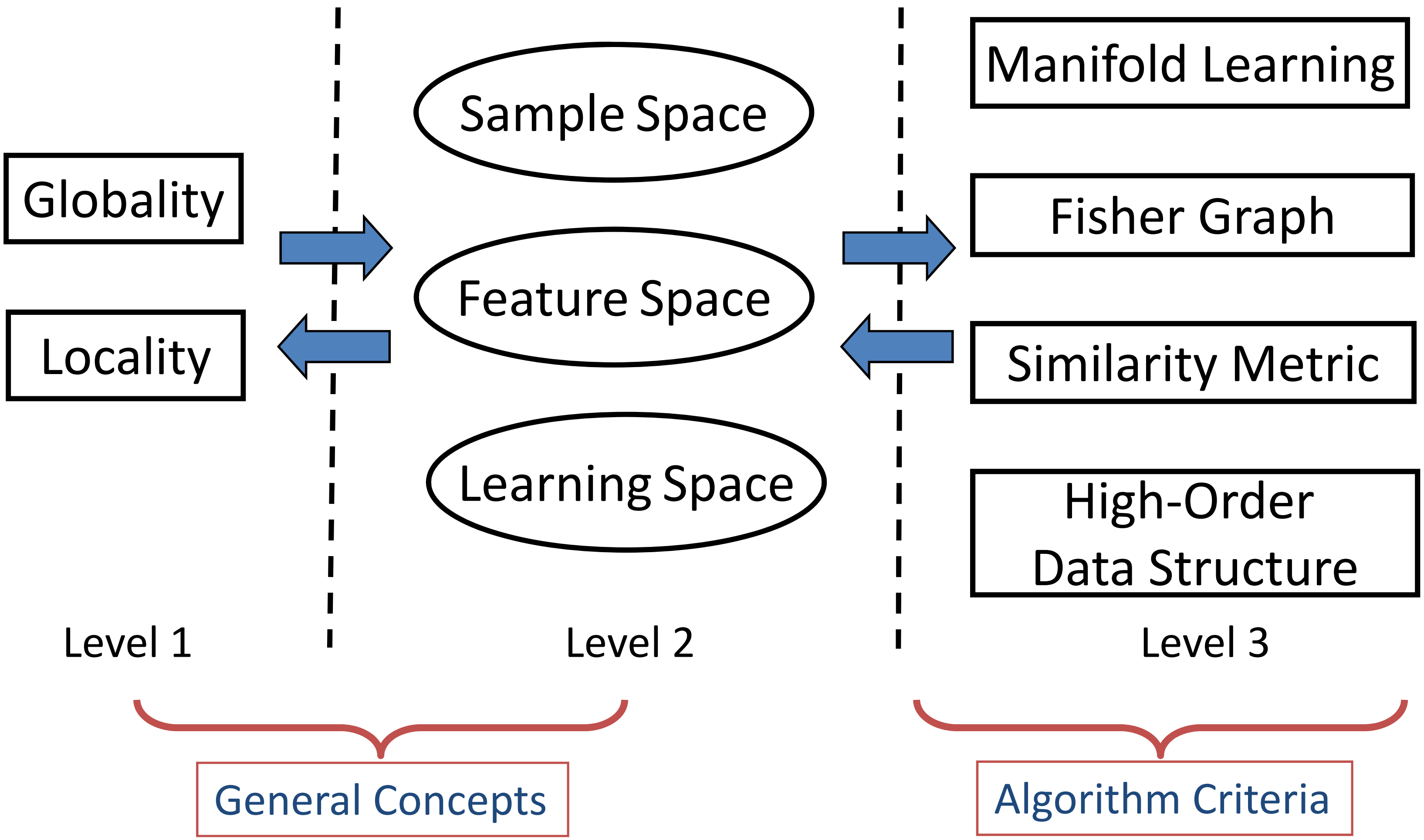
Facial Image Analysis

- Goal: Interpret given facial images in terms of facial attributes (Expression, Aging, Illumination, and Pose).
- Challenge
  - Dimensionality redundancy
  - Large size database
  - Unknown distribution
  - Large variations
  - Multimodality data
- Motivations and Proposed Work
  - A general framework of discriminative subspace learning.
  - Extracting discriminative features to boost the discriminating power.
  - Four novel algorithms are designed based on the framework.
  - Real-world applications (hard biometrics and soft biometrics): face recognition, facial expression analysis, age estimation, head pose estimation, and lipreading.

Background: Existing Methods

- **Global and Local Learning Methods**
  - Local Learning vs. Global Learning, K. Huang, H. Yang, I. King, and M. R. Lyu; Global Versus Local Methods in Nonlinear Dimensionality Reduction, V. de Silva and J. Tenenbaum; Generalized principal component analysis (GPCA), Y. Ma, et. al.; Globally-Coordinated Locally-Linear Modeling, C.-B. Liu.
- **Localized Subspace Learning Methods**
  - Locally Embedded Linear Subspaces, Z. Li, L. Gao, and A. K. Katsaggelos; Locally Adaptive Subspace, Y. Fu, Z. Li, T.S. Huang, A.K. Katsaggelos.
- **Patches/Parts Based Methods**
  - Flexible X-Y Patches, M. Liu, S.C. Yan, Y. Fu, and T. S. Huang; Patch-based Image Correlation, G-D. Guo and C. Dyer.
- **Feature Extraction Methods**
  - Local Binary Pattern (LBP), T. Ojala, M. Pietikainen, and T. Maenpaa; Histogram of Oriented Gradient descriptor (HOG), N. Dalai and B. Triggs.
- **Nonlinear Graph Embedding Methods**
  - Locally Linear Embedding (LLE), S.T. Roweis & L.K. Saul; Isomap, J.B. Tenenbaum, V.de Silva, J.C. Langford; Laplacian Eigenmaps (LE), M. Belkin & P. Niyogi
- **Linear Subspace Learning Methods**
  - Principal Component Analysis (PCA), M.A. Turk & A.P. Pentland; Multidimensional Scaling (MDS), T.F. Cox and M.A.A. Cox; Locality Preserving Projections (LPP), X.F. He, S.C. Yan, Y.X. Hu
- **Fisher Graph Methods**
  - Linear Discriminant Analysis (LDA), R.A. Fisher; Marginal Fisher Analysis (MFA), S.C. Yan, et al.; Local Discriminant Embedding (LDE), H.-T. Chen, et al.
- **Tensor Subspace Learning Methods**
  - Two-dimensional PCA (TPCA), J. Yang, et.al.; Two-dimensional LDA (TLDA), J. Ye, et.al.; Tensor subspace analysis (TSA), X. He, et al.; Tensor LDE (TLDE), J. Xia, et al.; Rank-r approximation, H. Wang.
- **Correlation-based Subspace Learnng Methods**
  - Discriminative Canonical Correlation (DCC), T.-K. Kim, et al.; Correlation Discriminant Analysis (CDA), Y. Ma, et al.

Unified Learning Framework



Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

Level 1 and 2: Concepts

- Feature-Globality (FG): FG takes each training image as a single feature with each pixel being a dimension of the feature vector/matrix.
- Feature-Locality (FL): FL selects local parts or local patches in the global feature space to build multiple models.
- Sample-Globality (SG): SG, like conventional methods, apply all training sample points to build the global model.
- Sample-Locality (SL): SL partitions the training sample space or searches local neighborhoods of a query to build linear model in local sample sets.
- Learning-Globality (LG): In a graph embedding view, LG constructs a globally connected graph to measure the data affinity for learning the low-dimensional representation.
- Learning-Locality (LL): LL constructs a partially connected graph embodying local connectivity.

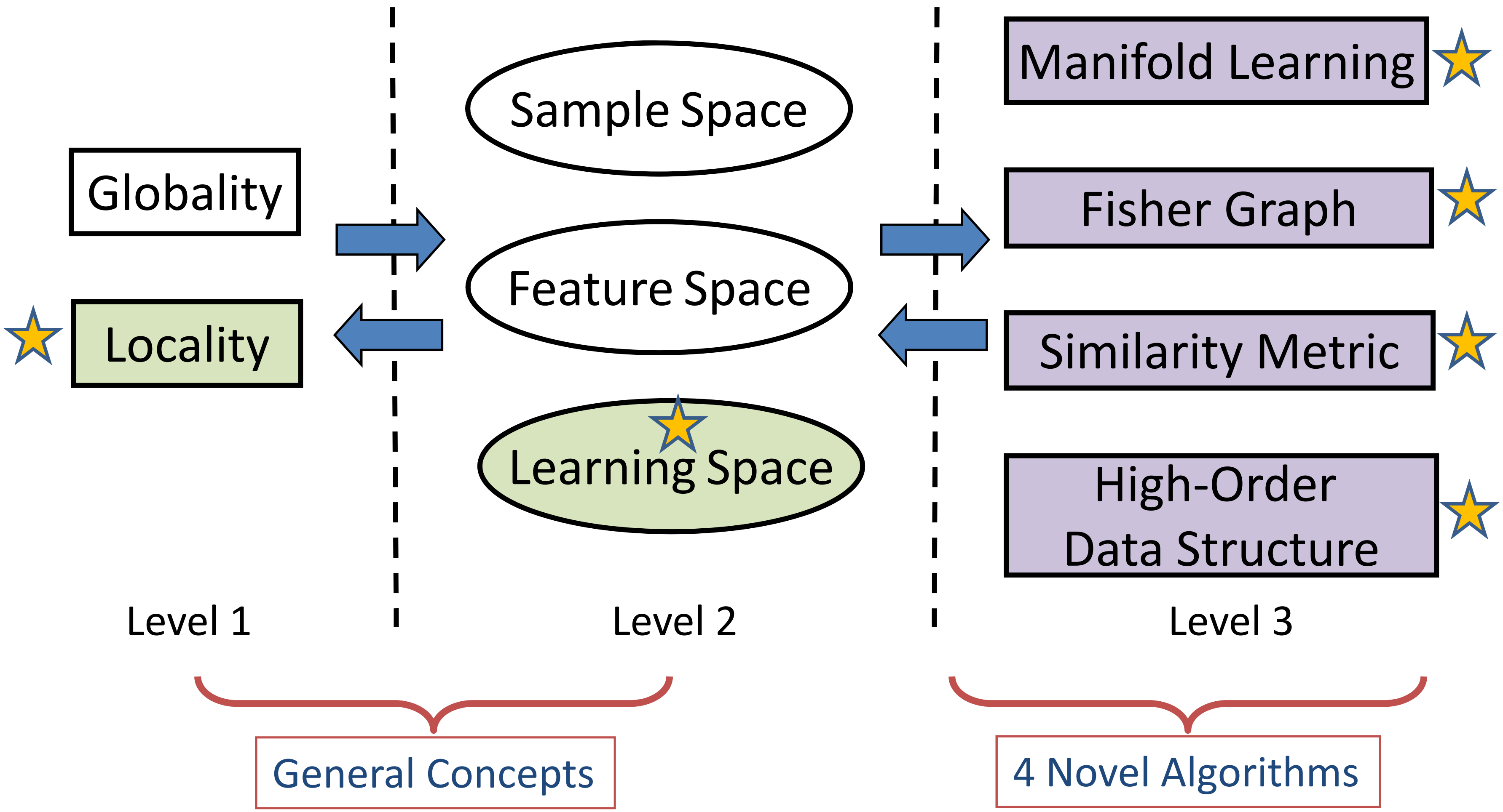
Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

Global vs. Local

- **Y. Fu**, Z. Li, J. Yuan, Y. Wu, and T. S. Huang, *Locality vs. Globality: Query-Driven Localized Linear Models for Facial Image Computing*, IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2008. (to appear soon)
- Discriminating Power Coefficient (DPC) is defined in terms of Graph Embedding theory.
- Locality manner outperforms the globality manner in many cases.
- Three locality concepts can be applied either individually or jointly.
- When should we choose locality instead of globality?
  - Small training sample case
    - Learning-Locality
    - Feature-Locality
  - Large training sample case
    - Sample-Locality + Feature-Locality
    - Sample-Locality + Learning-Locality
    - Sample-Locality + Feature-Locality + Learing-Locality

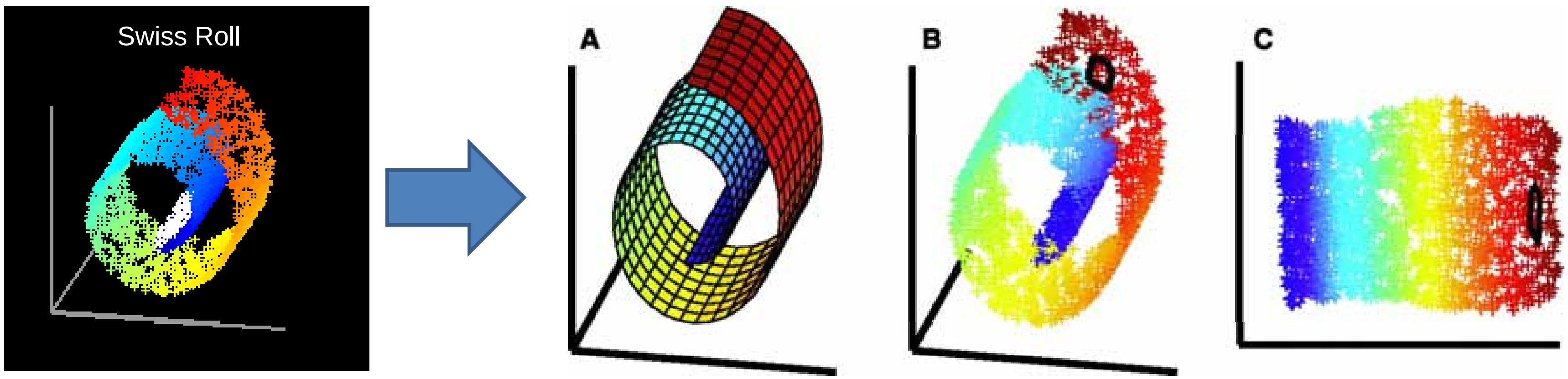
Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

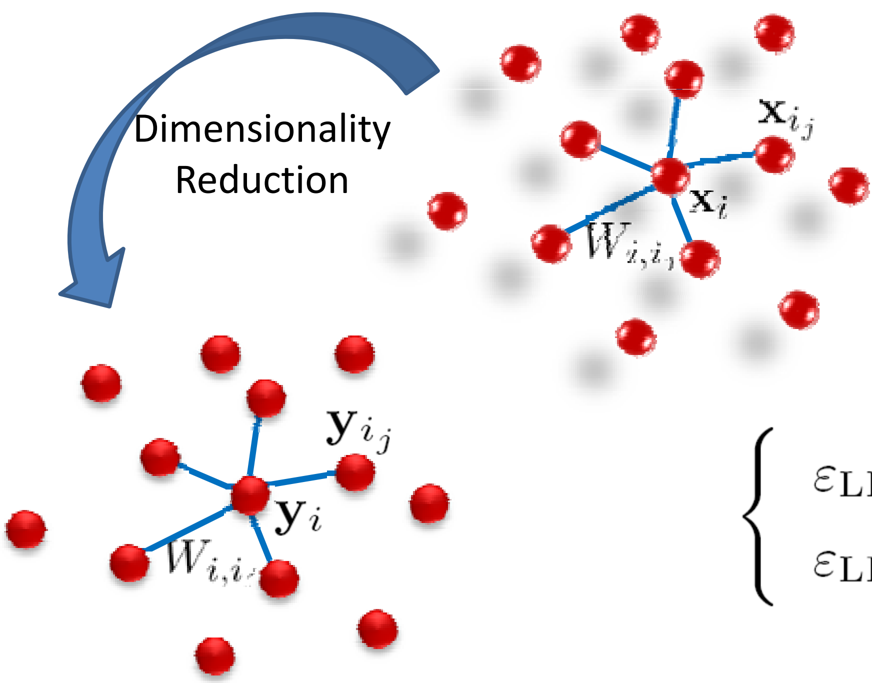
Unified Learning Framework



Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

Level 3: Manifold Learning





$$\varepsilon(W) = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{x}_{i_j} \right\|^2,$$

$$\begin{cases} \varepsilon_{\mathrm{LLE}} = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{y}_{i_j} \right\|^2 \\ \varepsilon_{\mathrm{LEA}} = \sum_{i=1}^{n} \left\| \mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{P}^T \mathbf{x}_{i_j} \right\|^2 \end{cases}$$

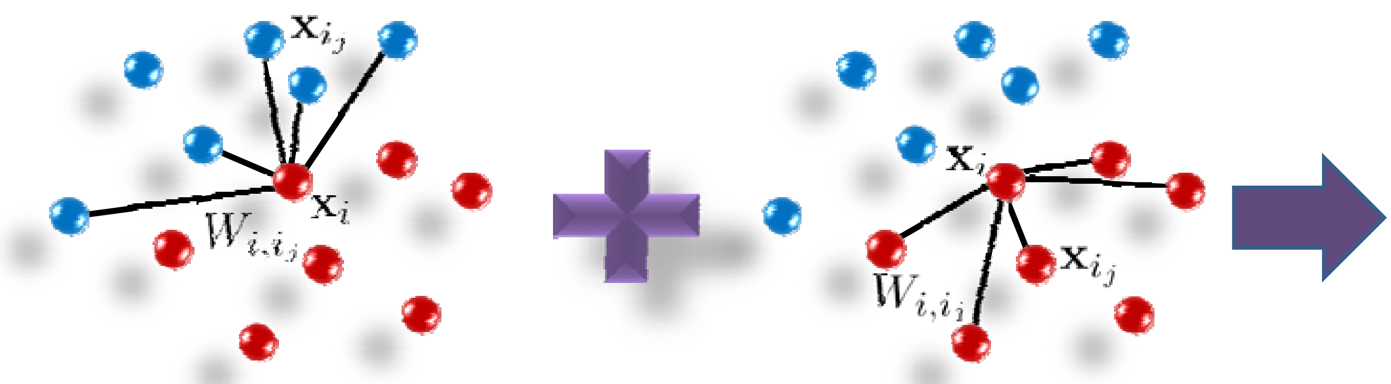
Courtesy of Sam T. Roweis and Lawrence K. Saul, Sience 2002

Level 3: Fisher Graph

- Graph Embedding (*S. Yan, IEEE TPAMI, 2007*)
  - *G*={*X*, *W*} is an undirected weighted graph.
  - *W* measures the similarity between a pair of vertices.
  - Laplacian matrix $L = D - W, \quad D_{ii} = \sum_{j \neq i} W_{ij}, \quad \forall\, i.$
  - Most manifold learning method can be reformulated as

$$y^* = \arg \min_{y^T B y = d} \sum_{i \neq j} \|y_i - y_j\|^2 W_{ij} = \arg \min_{y^T B y = d} y^T L y,$$

where *d* is a constant and *B* is the constraint matrix.



$$\begin{cases} \arg\min \mathbf{y}^T \mathbf{L}_{\mathrm{w}} \mathbf{y} \\ \arg\max \mathbf{y}^T \mathbf{L}_{\mathrm{b}} \mathbf{y} \end{cases}$$

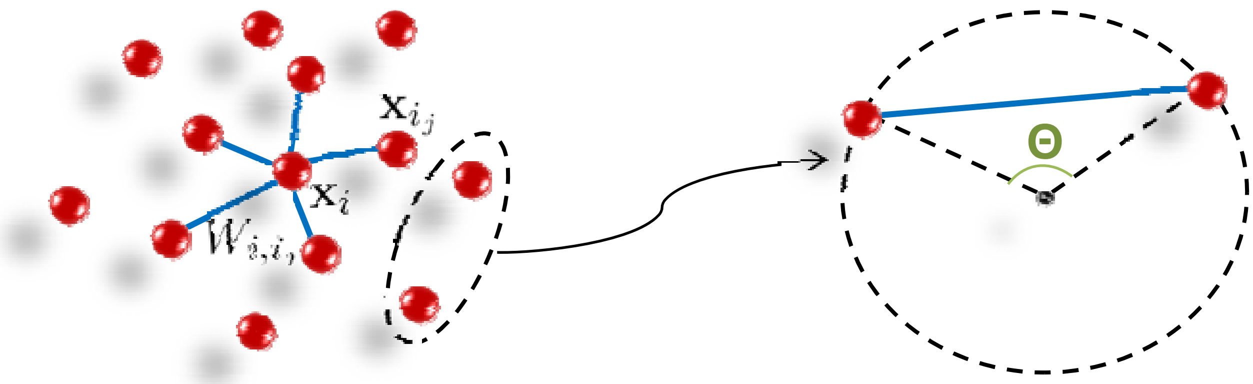
Between-Locality Graph

Within-Locality Graph

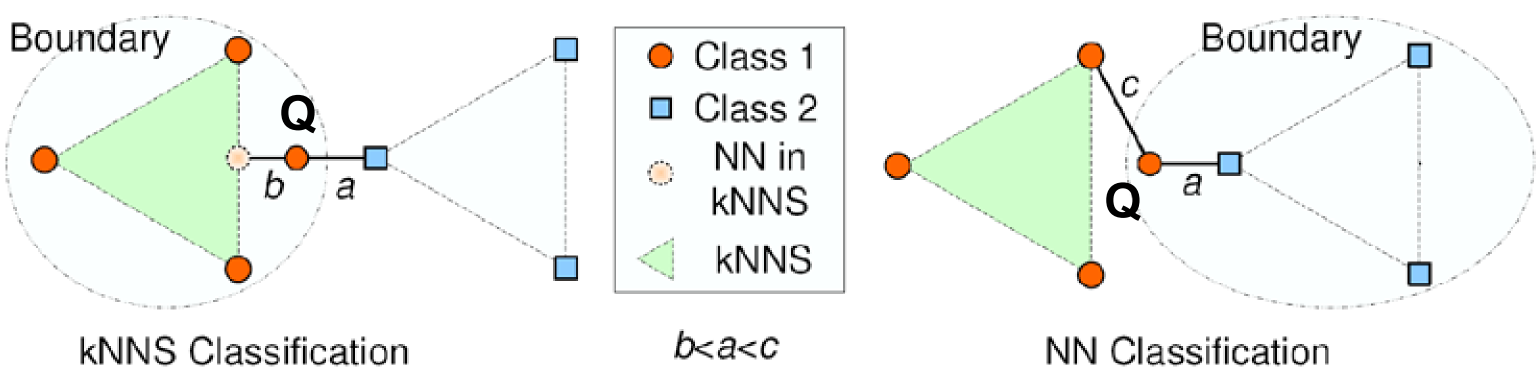
Courtesy of Shuicheng Yan

Level 3: Similarity Metric

- Single-Sample Metric
  - Euclidean Distance and Pearson Correlation Coefficient.

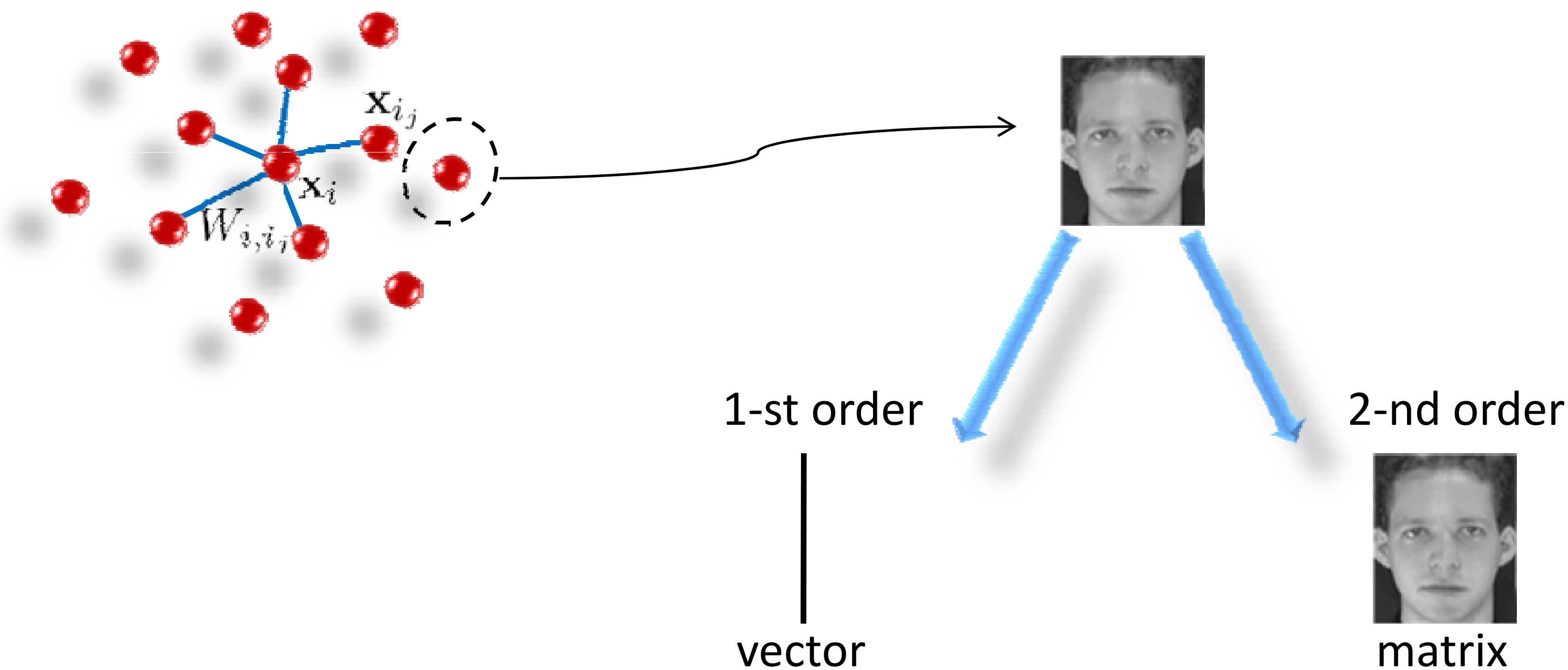


- Multi-Sample Metric
  - k-Nearest- Neighbor Simplex $S(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k) = \{ \sum_{i=1}^{k} \ell_i \mathbf{x}_i \mid \sum_{i=1}^{k} \ell_i = 1, \ell_i \geq 0 \}.$

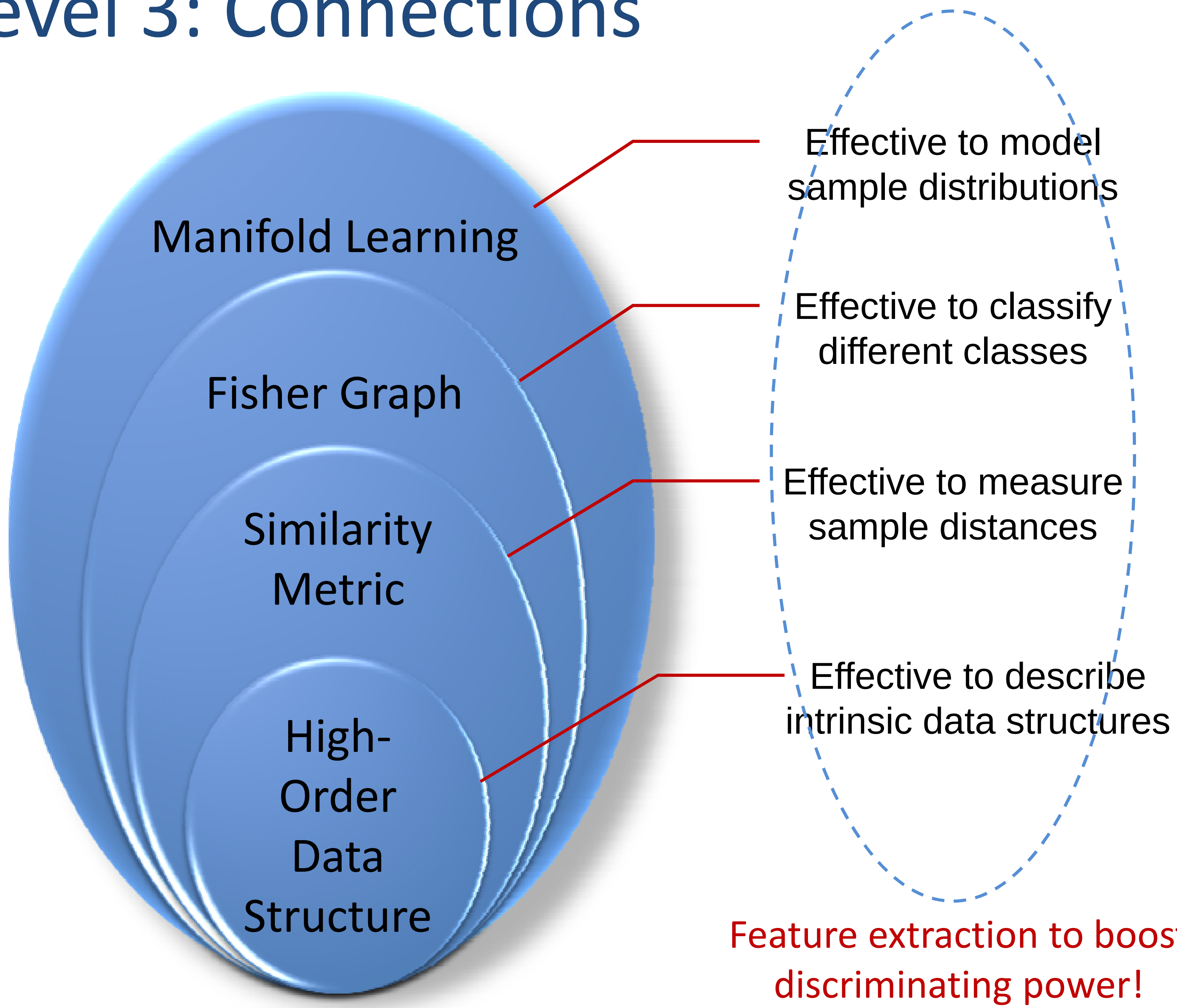


Level 3: High-Order Data Structure

- m-th order tensors $\{\mathbf{X}_i \mid \mathbf{X}_i \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}\}_{i=1}^{n}$
- Representation $\{\mathbf{Y}_i \mid \mathbf{Y}_i \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_m}\}_{i=1}^{n}$ where $d_i \leqslant D_i$
- Define $\mathbf{Y}_i = \mathbf{X}_i \times_1 \mathbf{U}_1 \times_2 \ldots \times_m \mathbf{U}_m$ where $\mathbf{U}_j \in \mathbb{R}^{D_j \times d_j}$
- Here, tensor means multilinear representation.

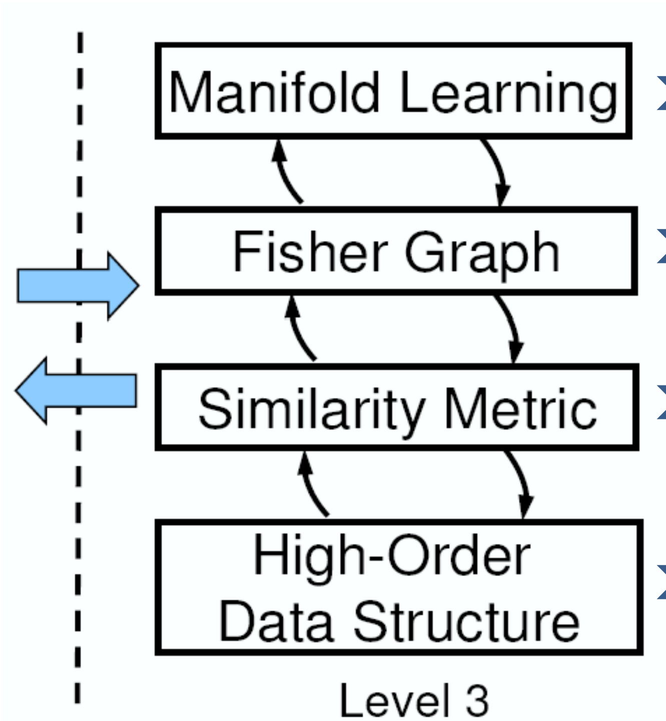


Level 3: Connections



Feature extraction to boost discriminating power!

Four Novel Algorithms



Locally Embedded Analysis (LEA)

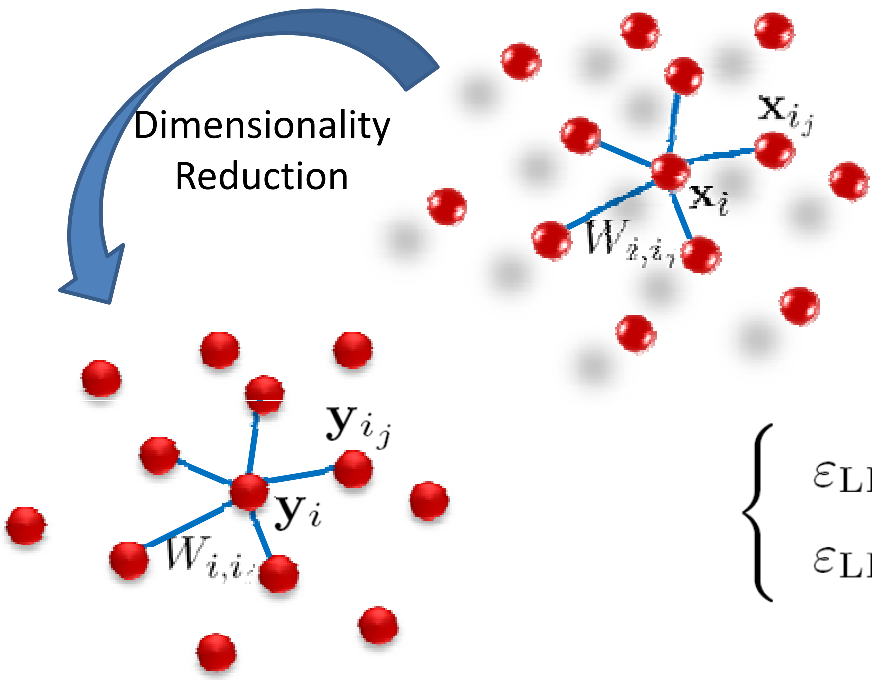
Discriminant Simplex Analysis (DSA)

Correlation Embedding Analysis (CEA)

Correlation Tensor Analysis (CTA)

Y. Fu, et. al., IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PMAI), 2008.
Y. Fu, et. al., IEEE Transactions on Image Processing (T-IP), 2008.
Y. Fu, et. al., IEEE Transactions on Multimedia (T-MM), 2008.
Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT), 2008.
Y. Fu, et. al., IEEE Transactions on Information Forensics and Security (T-IFS), 2008.
Y. Fu, et. al., Computer Vision and Image Understanding (CVIU), 2008.

1 Locally Embedded Analysis

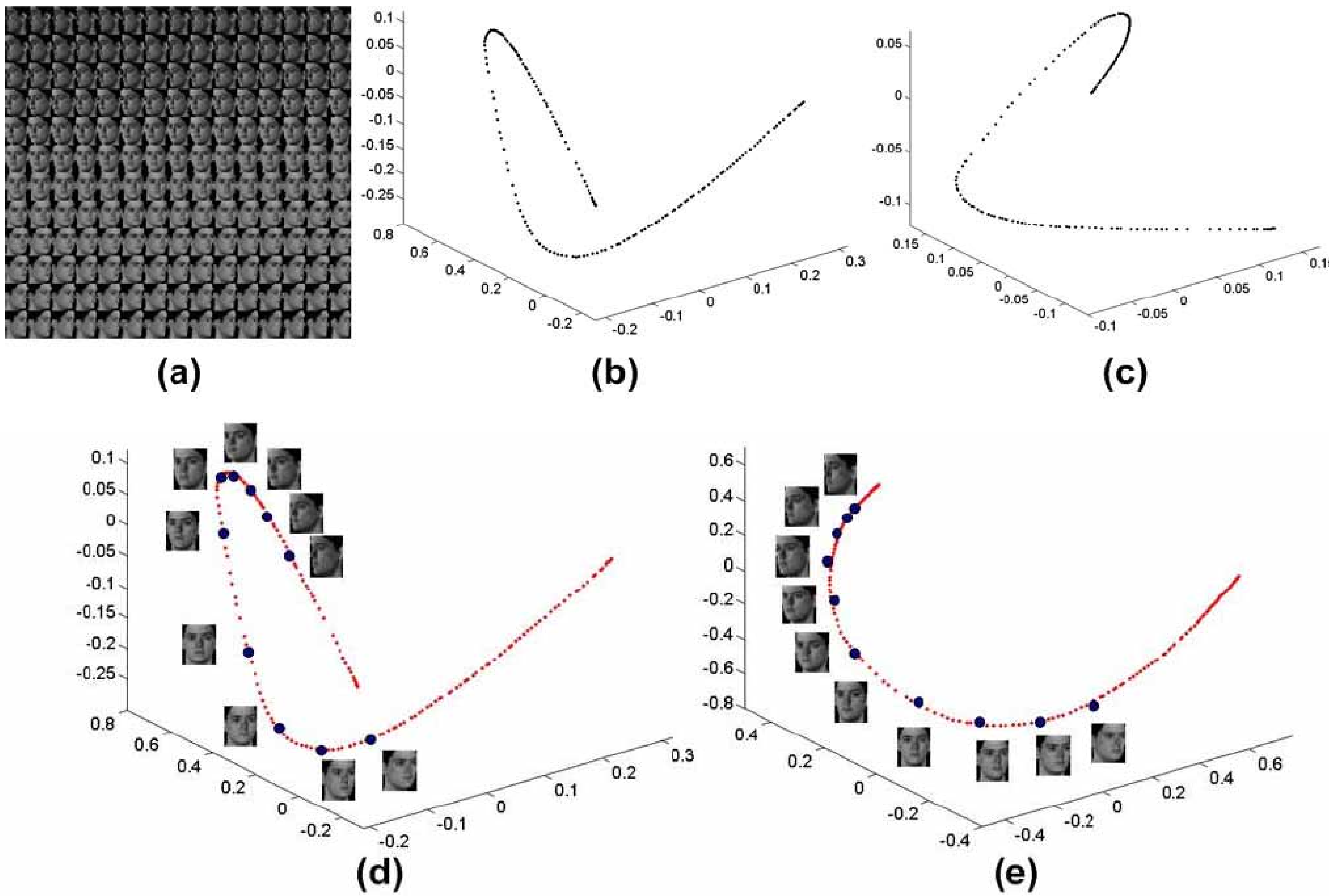


$$\varepsilon(W) = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{x}_{i_j} \right\|^2,$$

$$\begin{cases} \varepsilon_{\mathrm{LLE}} = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{y}_{i_j} \right\|^2 \\ \varepsilon_{\mathrm{LEA}} = \sum_{i=1}^{n} \left\| \mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^{k} W_{i,i_j} \mathbf{P}^T \mathbf{x}_{i_j} \right\|^2 \end{cases}$$

Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

LEA for Manifold Visualization



LEA for Face Recognition


ORL Database

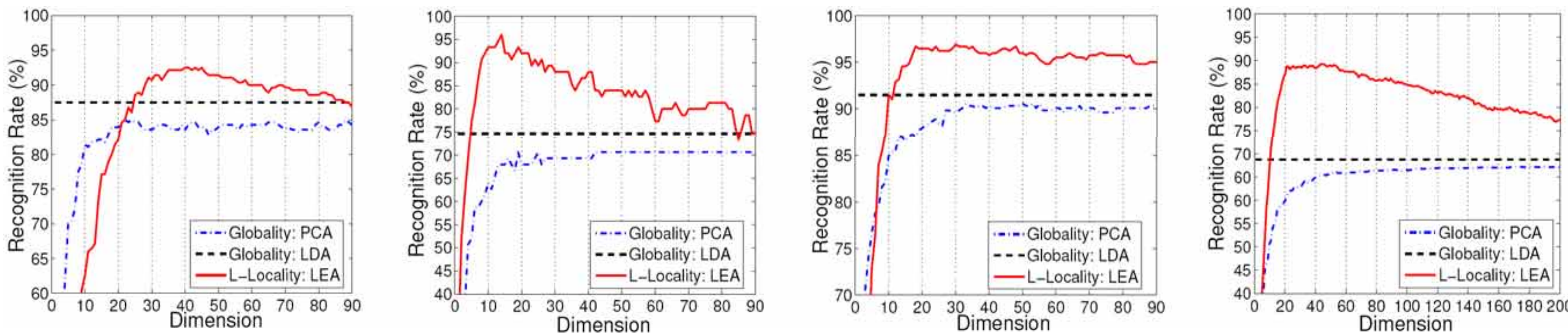

YALE Database


UMIST Database


UIUC-IFP-Y Internal Database

Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

LEA for Face Recognition



(a) ORL (b) YALE (c) UMIST (d) UIUC-IFP-Y

Method	ORL		YALE		UMIST		UIUC-IFP-Y	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
Unsupervised Globality	14.4%	88	25.8%	42	10.7%	33	32.7%	145
Supervised Globality	12.8%	39	23.6%	14	9.1%	19	31.2%	21
Learning-Locality	**9.7%**	40	**7.7%**	14	**4.3%**	18	**10.7%**	32

- 4 benchmark databases. 100-times running average.
- Randomly select 3, 6 and 25% images of each subject in ORL, YALE and UMIST for training and the rest for test.
- In the UIUC-IFP-Y, the 20, 10 and 130 images of each subject are randomly selected for training, gallery and test.

Y. Fu, et. al., IEEE Transactions on Circuits and Systems for Video Technology, 2008.

LEA for Face Pose Estimation

TABLE II

EVALUATION ON POINTING'04 HEAD POSE ESTIMATION WITH KNOWN NOSE-TIP LOCATIONS.

Metric vs. Method	LAA-NN [36]	Tensor	PCA	NPE/NPP	LEA
Pan error Mean	7.3°	6.2°	5.9°	6.0°	$\mathbf{5.5^\circ}$
Tilt error Mean	12.1°	8.6°	5.6°	5.1°	$\mathbf{4.8^\circ}$
Pan Class. Accu. ($\pm 0^\circ$)	61.3%	72.4%	66.4%	70.9%	**73.1%**
Tilt Class. Accu.($\pm 0^\circ$)	53.8%	75.7%	61.7%	76.8%	**77.7%**

TABLE IV

EVALUATION ON POINTING'04 HEAD POSE ESTIMATION WITH UNKNOWN NOSE-TIP LOCATIONS

Metric vs. Method	LAA-NN [36]	Tensor	PCA	NPE/NPP	LEA
Pan error Mean	10.3°	12.9°	14.8°	10.7°	$\mathbf{10.1^\circ}$
Tilt error Mean	15.9°	17.9°	14.5°	13.8°	$\mathbf{11.4^\circ}$
Pan Class. Accu. ($\pm 0^\circ$)	50.4%	49.3%	49.4%	50.5%	**53.1%**
Tilt Class. Accu. ($\pm 0^\circ$)	43.9%	54.8%	55.6%	55.8%	**58.4%**

TABLE V

EVALUATION ON CLEAR'07 CHIL HEAD POSE ESTIMATION

Metric vs. Method	Bayesian[39]	NN[40]	PCA	NPE/NPP	LEA
Pan error Mean	29.52°	8.5°	9.54°	9.22°	$\mathbf{8.13^\circ}$
Tilt error Mean	16.32°	12.5°	15.43°	11.24°	$\mathbf{9.86^\circ}$
Roll error Mean	35.67°	16.4°	5.46°	5.11°	$\mathbf{4.34^\circ}$
Average error Mean	27.17°	12.47°	10.14°	8.52°	$\mathbf{7.43^\circ}$

Y. Fu, et. al., IEEE Transactions on Information Forensics and Security, 2008.

LEA for Facial Expression Clustering

Table 3.3: LEC experimental results and algorithm comparison on Frey's face images.

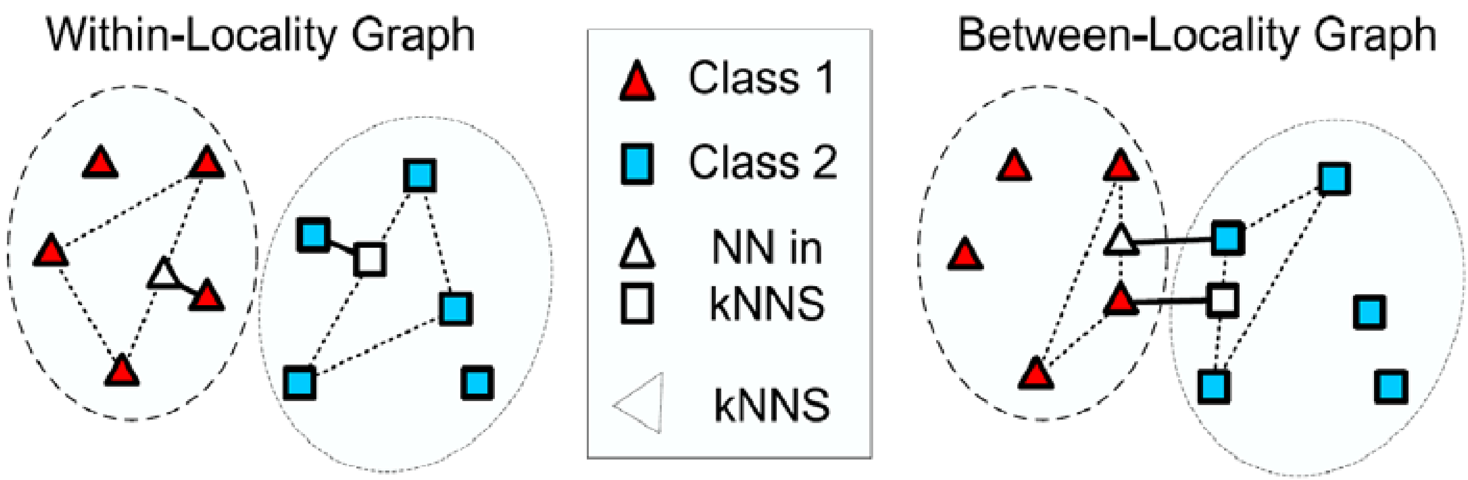
Method	**Clusters** $\sharp$ (c)	**K-means Replicates**	**Parameter** (r, σ^2)	**Dim.** $\sharp$ **(T., B.)**	**Error Rate** (%) ($\sharp$ / 1839)
K-means	3	50	None	336	15.72 (289/1839)
PCA+K-means	3	50	None	10 (T)	13.32 (245/1839)
PCA+K-means	3	50	None	1 (T)	15.39 (283/1839)
K-W+K-means	3	50	$\sigma^2 = 0.5$	10 (T)	9.84 (181/1839)
NPE+K-means	3	50	$r = 0.6$	3 (B)	8.59 (158/1839)
NJW	3	50	$\sigma^2 = 0.15$	3 (T)	7.34 (135/1839)
LEC-2	3	50	$r = 0.6$	1 (B)	**4.73 (87/1839)**

Table 3.4: LEC experimental results and algorithm comparison on AAI male face images.

Method	**Clusters** $\sharp$ (c)	**K-means Replicates**	**Parameter** (r, σ^2)	**Dim.** $\sharp$ **(T., B.)**	**Error Rate** (%) ($\sharp$ / 600)
K-means	3	50	None	780	40.83 (245/600)
PCA+K-means	3	50	None	100 (T)	40.83 (245/600)
PCA+K-means	3	50	None	1 (T)	35.33 (212/600)
K-W+K-means	3	50	$\sigma^2 = 0.5$	3 (T)	29.00 (174/600)
NPE+K-means	3	50	$r = 0.005$	3 (B)	33.17 (199/600)
NJW	3	50	$\sigma^2 = 0.1$	3 (T)	30.67 (184/600)
LEC-2	3	50	$r = 0.005$	1 (B)	**20.67 (124/600)**

Y. Fu, et. al., Computer Vision and Image Understanding, 2008.

2 Discriminant Simplex Analysis



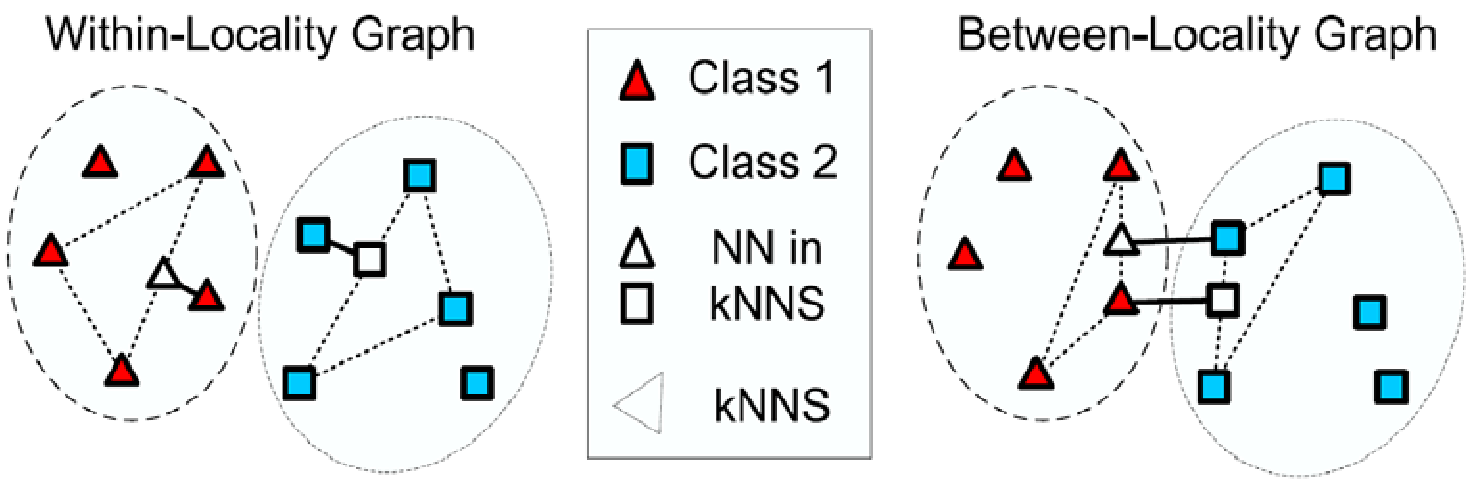
$$\begin{cases} \varepsilon_{\mathrm{w}}(\mathbf{C}_{\mathrm{w}}) = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{k_{\mathrm{w}}} c_{\mathrm{w}(j)}^{(i)} \mathbf{x}_{\mathrm{w}(j)}^{(i)} \right\|^2 \\ \varepsilon_{\mathrm{b}}(\mathbf{C}_{\mathrm{b}}) = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{k_{\mathrm{b}}} c_{\mathrm{b}(j,l^*)}^{(i)} \mathbf{x}_{\mathrm{b}(j,l^*)}^{(i)} \right\|^2 \end{cases}$$

$$\begin{cases} \varepsilon_{\mathrm{w}}(\mathbf{y}) = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{j=1}^{k_{\mathrm{w}}} c_{\mathrm{w}(j)}^{(i)} \mathbf{y}_{\mathrm{w}(j)}^{(i)} \right\|^2 \\ \varepsilon_{\mathrm{b}}(\mathbf{y}) = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{j=1}^{k_{\mathrm{b}}} c_{\mathrm{b}(j,l^*)}^{(i)} \mathbf{y}_{\mathrm{b}(j,l^*)}^{(i)} \right\|^2 \end{cases}$$

$l^* \in \{1, 2, \ldots, c-1\}$ indicates the label of the nearest different class.

Y. Fu, et. al., IEEE Transactions on Information Forensics and Security, 2008.

2 Discriminant Simplex Analysis



$$\mathbf{x} \Rightarrow \tilde{\mathbf{y}} = \arg\max \frac{\varepsilon_{\mathrm{b}}(\mathbf{y})}{\varepsilon_{\mathrm{w}}(\mathbf{y})} = \arg\max \frac{\mathbf{y}^T \mathbf{L}_{\mathrm{b}} \mathbf{y}}{\mathbf{y}^T \mathbf{L}_{\mathrm{w}} \mathbf{y}}.$$

$$\begin{cases} \varepsilon_{\mathrm{w}}(\mathbf{P}) = \sum_{i=1}^{n} \left\| \mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^{k_{\mathrm{w}}} c_{\mathrm{w}(j)}^{(i)} \mathbf{P}^T \mathbf{x}_{\mathrm{w}(j)}^{(i)} \right\|^2 \\ \varepsilon_{\mathrm{b}}(\mathbf{P}) = \sum_{i=1}^{n} \left\| \mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^{k_{\mathrm{b}}} c_{\mathrm{b}(j,l^*)}^{(i)} \mathbf{P}^T \mathbf{x}_{\mathrm{b}(j,l^*)}^{(i)} \right\|^2 \end{cases}$$

$$\mathbf{X}\mathbf{L}_{\mathrm{b}}\mathbf{X}^T\mathbf{P} = \mathbf{X}\mathbf{L}_{\mathrm{w}}\mathbf{X}^T\mathbf{P}\Lambda$$

Y. Fu, et. al., IEEE Transactions on Information Forensics and Security, 2008.

DSA for Face Recognition

ORL Database

PIE Database

- *ORL face database*: 400 images of 40 subjects. Size of 32x32.
- *CMU PIE database*: 41368 images of 68 subjects. Randomly select 34 images per individual. Size of 20x20.

FACE RECOGNITION ACCURACY COMPARISON ON THE ORL DATABASE USING THE MULTISAMPLE METRIC

Method Accuracy	3-Train (%)[Dim.]	4-Train (%)[Dim.]	5-Train (%)[Dim.]
NFL + NN	76.79 [117]	79.58 [58]	84.00 [133]
2NNS + NN	78.21 [107]	82.5 [106]	87.5 [158]
kNNS + NN	**81.07** [119]	**86.25** [84]	**89.50** [111]

FACE RECOGNITION ACCURACY COMPARISON ON THE ORL DATABASE

Method Accuracy	3-Train (%)[Dim.]	4-Train (%)[Dim.]	5-Train (%)[Dim.]
Baseline	76.79 [1024]	80.00 [1024]	84.00 [1024]
Eigenface	76.79 [63]	80.00 [64]	84.00 [133]
Fisherface	86.07 [39]	90.00 [38]	95.00 [34]
Laplacianface	86.43 [41]	92.92 [36]	96.00 [76]
DSAface	**90.71** [60]	**95.42** [20]	**97.00** [33]

TABLE IV
FACE RECOGNITION ACCURACY COMPARISON ON THE PIE DATABASE

Method Accuracy	5-Train (%)[Dim.]	10-Train (%)[Dim.]	15-Train (%)[Dim.]	20-Train (%)[Dim.]
Baseline	24.34 [400]	36.46 [400]	45.20 [400]	52.84 [400]
Eigenface	24.34 [200]	36.46 [245]	45.20 [285]	52.84 [282]
Fisherface	51.27 [70]	66.42 [51]	73.22 [46]	78.89 [45]
Laplacianface	37.32 [196]	63.85 [67]	76.24 [59]	82.35 [66]
DSAface	**52.54** [141]	**67.28** [114]	**78.10** [51]	**83.72** [54]

Y. Fu, et. al., IEEE Transactions on Information Forensics and Security, 2008.

DSA for Lipreading

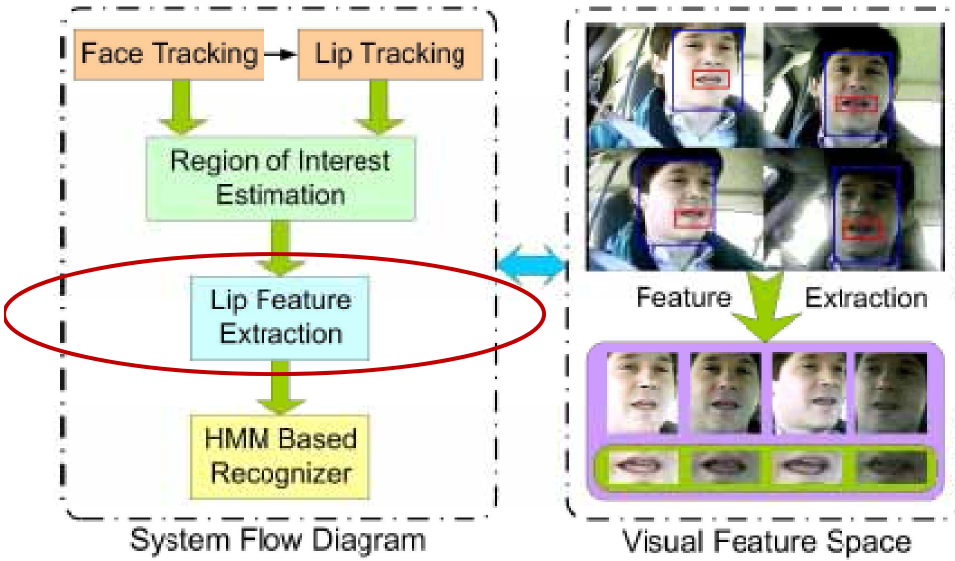
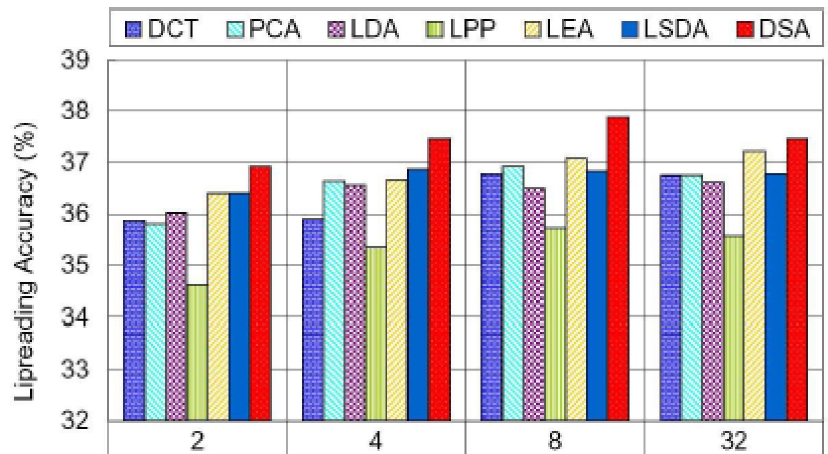


Figure 6. Lipreading system structure.



	2	4	8	32
DCT	35.87 (49)	35.90 (37)	36.76 (49)	36.73 (44)
PCA	35.80 (45)	36.63 (42)	36.91 (47)	36.73 (41)
LDA	36.02 (48)	36.55 (48)	36.48 (49)	36.60 (49)
LPP	34.61 (49)	35.34 (41)	35.72 (48)	35.57 (38)
LEA	36.38 (42)	36.65 (40)	37.08 (44)	37.21 (33)
LSDA	36.38 (48)	36.86 (49)	36.81 (39)	36.76 (47)
DSA	36.91 (41)	37.41 (47)	37.87 (41)	37.46 (50)

Number of Gaussian Mixtures

Figure 7. Comparison of the highest lipreading accuracy and dimensionality reduction of all the 7 methods under 4 different numbers of Gaussian mixtures per HMM state.

- Training: 21 talkers
- Test: 13 *different* talkers
- Dim: 584x4DCT→100PCA

This is the highest lipreading accuracy on this database ever reported.

Y. Fu, et. al., IEEE Transactions on Information Forensics and Security, 2008.

3 Correlation Embedding Analysis

Pearson Correlation Coefficient (PCC):

$$\mathrm{Corr}(X_1, X_2) = \frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{\sqrt{\langle \mathbf{x}_1, \mathbf{x}_1 \rangle}\sqrt{\langle \mathbf{x}_2, \mathbf{x}_2 \rangle}},$$

Two types of PCC: Centered PCC and Uncentered PCC.

- **Why Correlation?**
  - Correlation metric outperforms Euclidean distance in many cases for visual classification purpose.
  - CEA has the capability to handle data on a hypersphere, which cannot be well explained by conventional Euclidean distance based methods.

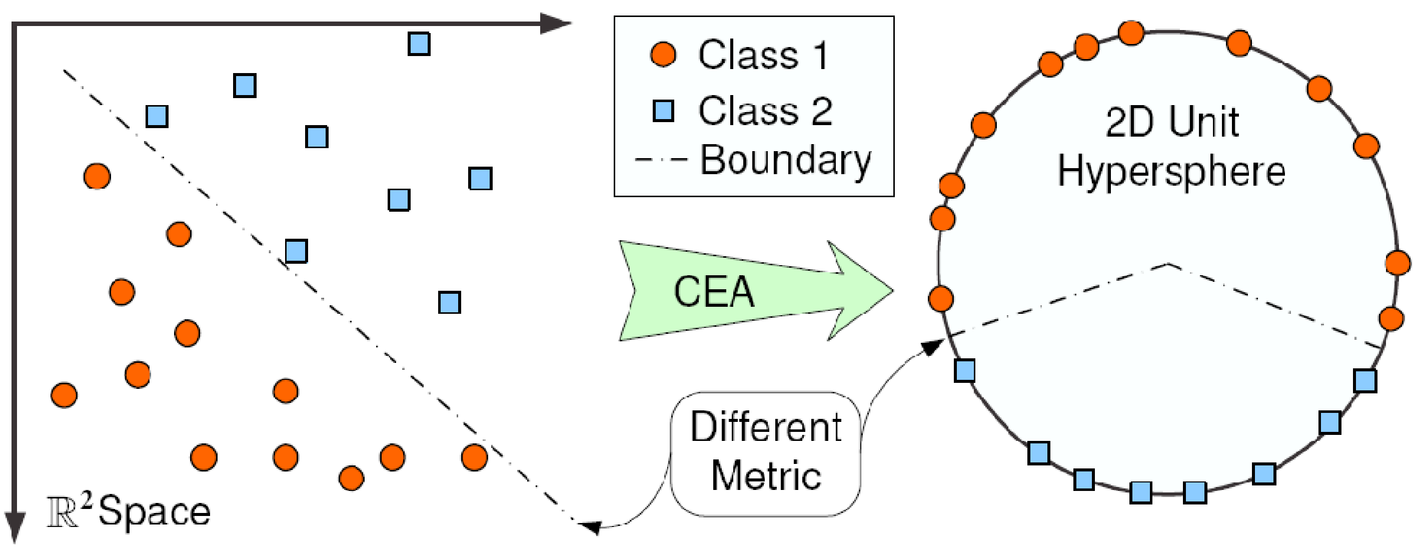
- M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein, *Cluster Analysis and Display of Genome-wide Expression Patterns*, Proc. of National Academy of Sciences of USA, 1998.
- B.V.K. Vijaya Kumar, R.D. Juday, and A. Mahalanobis, *Correlation Pattern Recognition*, Cambridge Uni. Press, 2006.
- T.-K. Kim, J. Kittler, and R. Cippola, *Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations*, IEEE Trans. on PAMI, 29(6):1005-1018, 2007.
- Yun Fu, Ming Liu, and Thomas S. Huang, *Conformal Embedding Analysis with Local Graph Modeling on the Unit Hypersphere*, IEEE CVPR'07-CA, 2007.
- Y. Ma, S. Lao, E. Takikawa, and M. Kawade, *Discriminant analysis in correlation similarity measure space*, Proc. of ICML, vol. 227, pp. 577-584, 2007.

Y. Fu, et. al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

3 Correlation Embedding Analysis

- Objective Function

$$\arg\max_{\mathbf{P}} \quad \varepsilon(\mathbf{P}) = \sum_{i,j=1}^{n} \underbrace{(1 - \frac{\mathbf{y}_i^T \mathbf{y}_j}{\| \mathbf{y}_i \| \| \mathbf{y}_j \|})}_{\text{Correlation Distance}} \cdot \underbrace{(w_{ij}^{(\mathrm{d})} - w_{ij}^{(\mathrm{s})})}_{\text{Fisher Graph}}$$



Y. Fu, et. al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

3 Correlation Embedding Analysis

- Learning CEA subspace

$$\arg\max_{\mathbf{P}} \quad \varepsilon(\mathbf{P}) = \sum_{i,j=1}^{n} \left(1 - \frac{\mathbf{x}_i \mathbf{P}\mathbf{P}^T \mathbf{x}_j}{\sqrt{(\mathbf{x}_i^T \mathbf{P}\mathbf{P}^T \mathbf{x}_i)(\mathbf{x}_j^T \mathbf{P}\mathbf{P}^T \mathbf{x}_j)}}\right) \cdot w_{ij}^{(\mathrm{d-s})}$$

$$\text{where} \quad \mathbf{y}_i = \mathbf{P}^T \mathbf{x}_i \quad \mathbf{P} = [\mathbf{p}_1 \mathbf{p}_2 \ldots \mathbf{p}_d]$$

$$(w_{ij}^{(\mathrm{d})} - w_{ij}^{(\mathrm{s})})$$

- Nonlinear, no closed-form solution
- Gradient decent rule

$$\frac{\partial \varepsilon(\mathbf{P})}{\partial \mathbf{P}} = \sum_{i,j=1}^{n} \left[\frac{f_{ij}(\mathbf{P}^T \mathbf{x}_i \mathbf{x}_i^T)_:}{f_i^3 f_j} + \frac{f_{ij}(\mathbf{P}^T \mathbf{x}_j \mathbf{x}_j^T)_:}{f_j^3 f_i} - \frac{(\mathbf{P}^T \mathbf{x}_i \mathbf{x}_j^T)_: + (\mathbf{P}^T \mathbf{x}_j \mathbf{x}_i^T)_:}{f_i f_j}\right] \cdot w_{ij}^{(\mathrm{d-s})},$$

$$f_i = \| \mathbf{y}_i \| = \sqrt{\mathbf{x}_i^T \mathbf{P}\mathbf{P}^T \mathbf{x}_i}, \quad f_j = \| \mathbf{y}_j \| = \sqrt{\mathbf{x}_j^T \mathbf{P}\mathbf{P}^T \mathbf{x}_j}, \quad f_{ij} = \mathbf{x}_i^T \mathbf{P}\mathbf{P}^T \mathbf{x}_j,$$

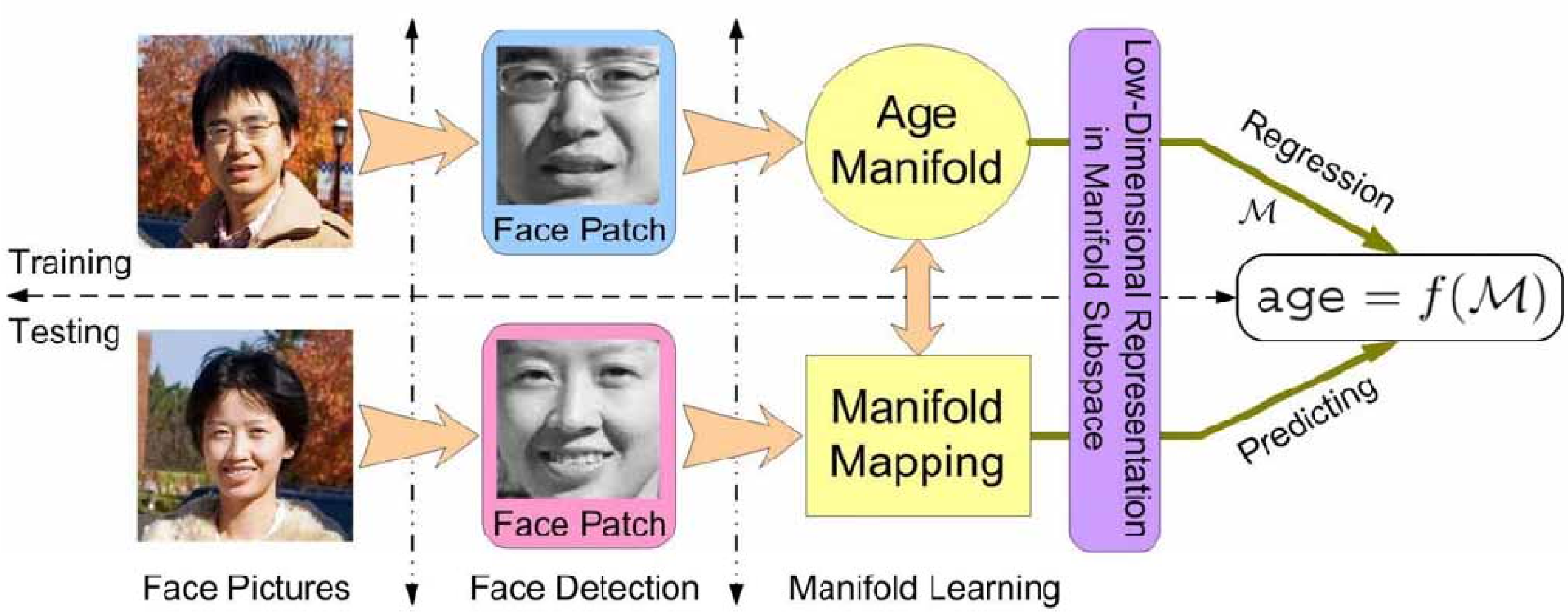
Y. Fu, et. al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

CEA for Face Recognition

FACE RECOGNITION PERFORMANCE COMPARISON ON THE YALE-B DATABASE.

Method	5 Train. Sample		10 Train. Sample		20 Train. Sample		30 Train. Sample	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
Baseline	54.73 %	1024	36.06%	1024	31.22%	1024	27.71%	1024
PCA	54.73 %	185	36.06%	367	31.22%	321	27.71%	377
CPCA	41.39 %	189	19.64 %	329	15.07 %	364	13.39 %	361
LDA	36.13%	37	19.40%	41	25.96%	51	21.67%	38
LPP	34.08%	87	18.03%	75	30.26%	211	20.20%	122
LDE	35.46%	49	19.49%	35	26.26%	38	29.33%	96
N-LDA	29.44%	43	9.75%	36	12.14%	37	14.86%	99
N-LPP	27.70%	75	9.80%	206	11.66%	153	16.87%	297
N-LDE	28.86%	61	10.62%	38	11.54%	44	13.85%	40
LDA-C	29.97%	48	13.40%	55	16.81%	59	20.90%	37
LPP-C	27.83%	97	13.94%	148	12.78%	151	23.80%	200
LDE-C	36.80%	176	20.08%	199	20.45%	190	15.87%	158
N-LDA-C	27.65%	90	9.80%	37	11.84%	37	14.40%	99
N-LPP-C	23.87%	189	8.38%	197	9.93%	180	13.16%	194
N-LDE-C	27.07%	64	9.41%	45	11.84%	37	13.93%	39
CEA	**22.61%**	103	**7.55%**	229	**6.22%**	218	**3.56%**	202

CEA for Age Estimation



Multiple linear regression $\mathbf{L} = \widetilde{\mathbf{Y}}\mathbf{B} + \mathbf{e}, \quad \mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{I},$

Model fitting $\hat{\mathbf{L}} = \widetilde{\mathbf{Y}}\hat{\mathbf{B}}.$ Ordinary Least Squares $\hat{\mathbf{B}} = (\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-1}\widetilde{\mathbf{Y}}'\mathbf{L} = \mathbf{HL}$

Residuals $\hat{\mathbf{e}} = \mathbf{L} - \hat{\mathbf{L}}.$ $\mathrm{E}(\hat{\mathbf{e}}) = 0$ $\mathrm{Var}(\hat{\mathbf{e}}) = \sigma^2(\mathbf{I} - \mathbf{H}).$

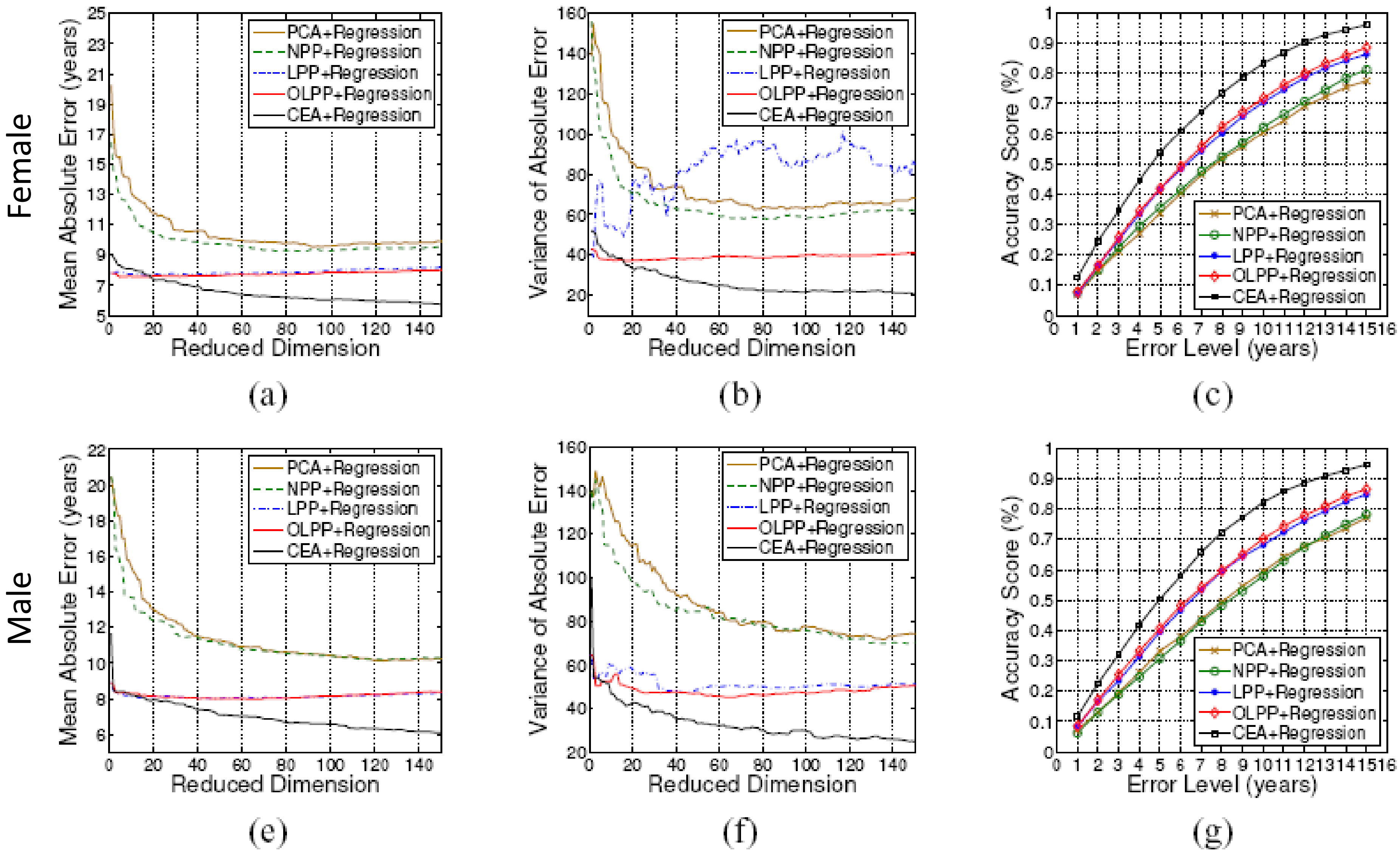
Quadratic function $\hat{l}_i = \hat{b}_0 + \hat{\mathbf{b}}_1^T\mathbf{y}_i + \hat{\mathbf{b}}_2^T\mathbf{y}_i^2,$

$$\hat{\mathbf{L}} = [\hat{l}_1 \cdots \hat{l}_n]^T, \quad \hat{\mathbf{B}} = [\hat{b}_0 \;\; \hat{b}_1^{(1)} \cdots \hat{b}_1^{(d)} \;\; \hat{b}_2^{(1)} \cdots \hat{b}_2^{(d)}]^T$$

$$\widetilde{\mathbf{Y}} = [\mathbf{1}_{n\times 1} \;\; [\mathbf{y}_1 \cdots \mathbf{y}_n]^T \;\; [\mathbf{y}_1^2 \cdots \mathbf{y}_n^2]^T].$$

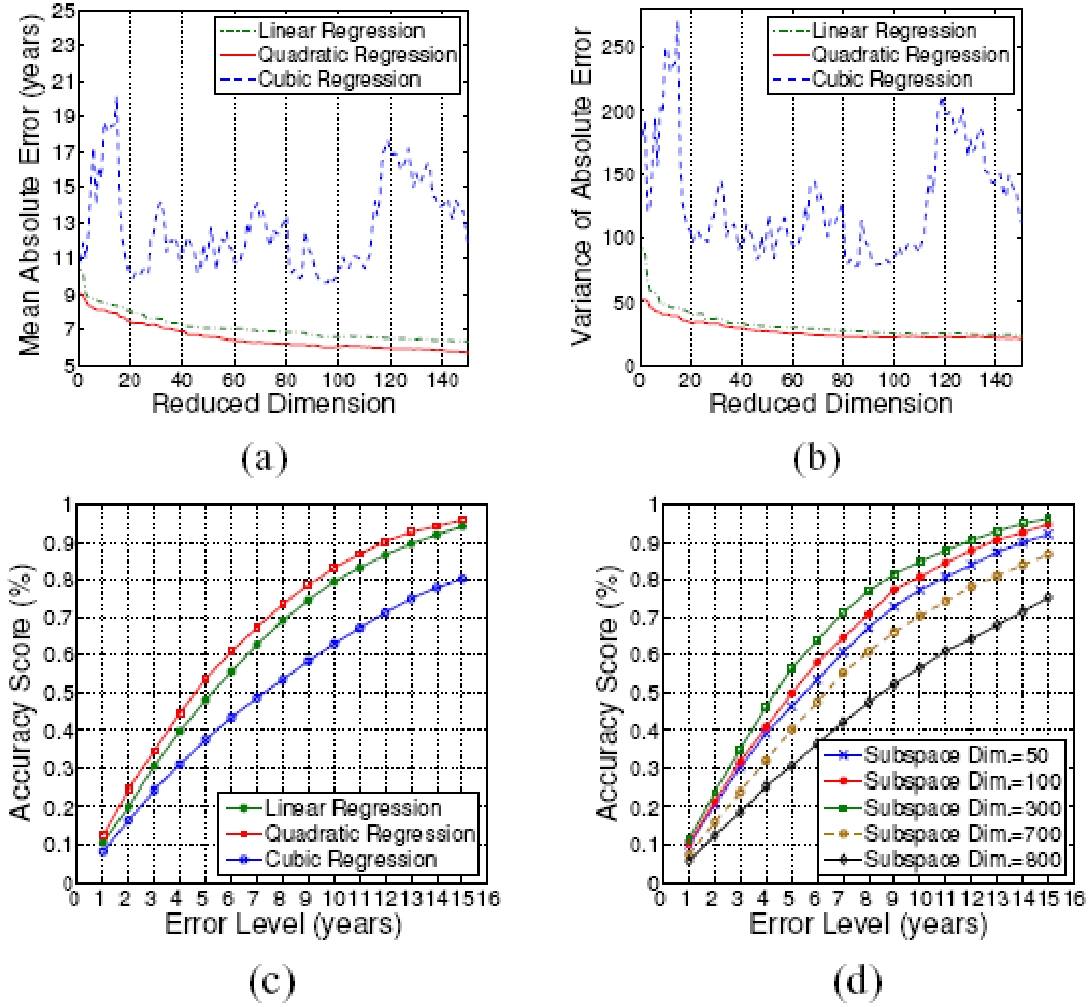
Y. Fu, et. al., IEEE Transactions on Multimedia, 2008.

CEA for Age Estimation



Y. Fu, et. al., IEEE Transactions on Multimedia, 2008.

CEA for Age Estimation



The result is also demonstrated by *A. Lanitis, et al, IEEE Trans. on SMC-B, 2004*.

Y. Fu, et. al., IEEE Transactions on Multimedia, 2008.

4 Correlation Tensor Analysis

Given two m-th order tensors, $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}$

Pearson Correlation Coefficient (PCC):

$$\mathrm{Corr}(X_1, X_2) = \frac{\langle \mathbf{X}_1, \mathbf{X}_2 \rangle}{\sqrt{\langle \mathbf{X}_1, \mathbf{X}_1 \rangle}\sqrt{\langle \mathbf{X}_2, \mathbf{X}_2 \rangle}},$$

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle = \textstyle\sum_{i_1=1, i_2=1, \ldots, i_m=1}^{D_1, D_2, \ldots, D_m} \mathbf{X}_{1; i_1, i_2, \ldots, i_m} \mathbf{X}_{2; i_1, i_2,}$$

CTA objective function

$$\arg\max_{\mathbf{U}} \ \varepsilon(\mathbf{U}) = \sum_{i,j=1}^{n} \underbrace{\left(1 - \frac{\langle \mathbf{Y}_i, \mathbf{Y}_j \rangle}{\sqrt{\langle \mathbf{Y}_i, \mathbf{Y}_i \rangle \langle \mathbf{Y}_j, \mathbf{Y}_j \rangle}}\right)}_{\text{Correlation Distance and Multilinear Representation}} \cdot \underbrace{\left(w_{ij}^{(\mathrm{d})} - w_{ij}^{(\mathrm{s})}\right)}_{\text{Fisher Graph}}$$

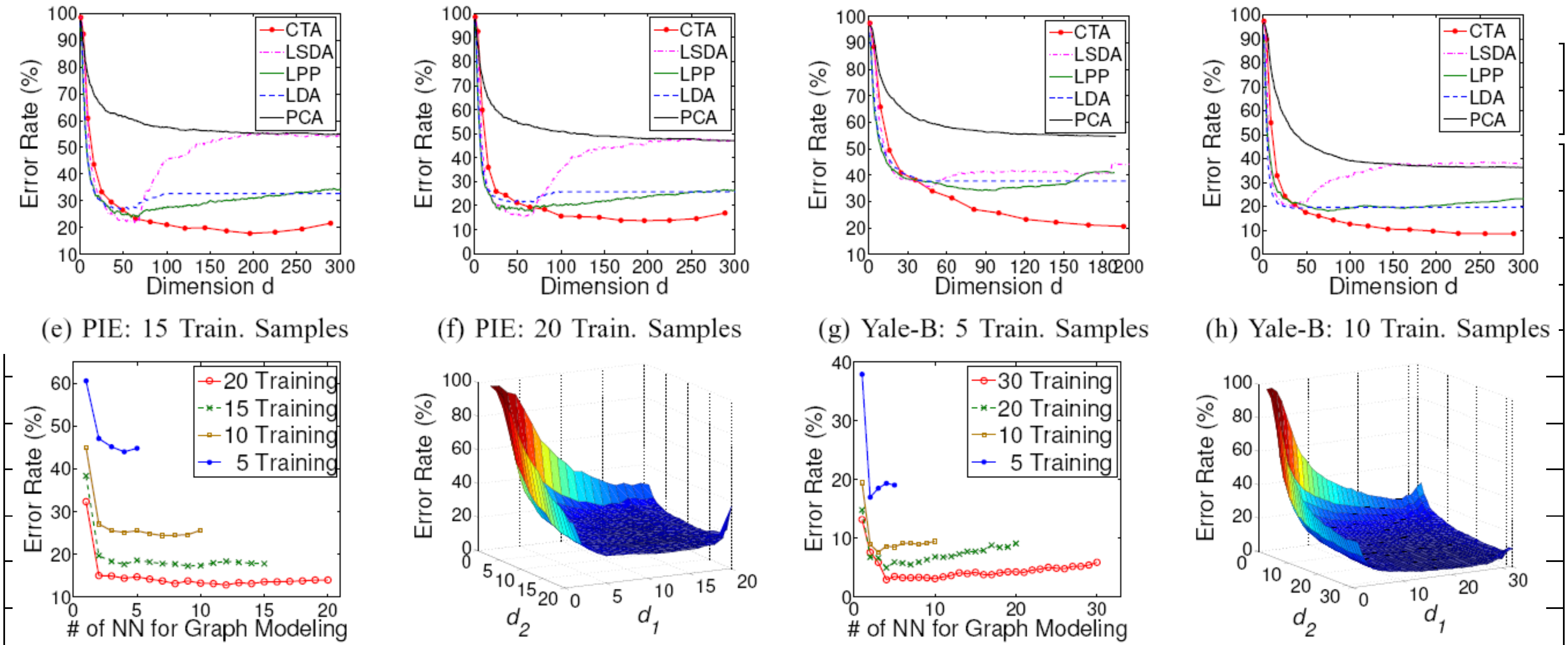
$$\mathbf{Y}_i = \mathbf{X}_i \times_1 \underbrace{\mathbf{U}_1 \times_2 \ldots \times_m \mathbf{U}_m}_{\text{m different subspaces}}, \ \mathbf{U}_j \in \mathbb{R}^{D_j \times d_j} \ \text{for } j = 1, \ldots, m.$$

Y. Fu, et. al., IEEE Transactions on Image Processing, 2008.

CTA for Face Recognition

TABLE II

FACE RECOGNITION PERFORMANCE COMPARISON ON THE YALE-B DATABASE.



(e) PIE: 15 Train. Samples (f) PIE: 20 Train. Samples (g) Yale-B: 5 Train. Samples (h) Yale-B: 10 Train. Samples

(a) PIE: Error vs. k (b) PIE: 20 Train. Samples (c) Yale-B: Error vs. k (d) Yale-B: 30 Train. Samples

Y. Fu, et. al., IEEE Transactions on Image Processing, 2008.

Discussion 1

- LEA can be both supervised and unsupervised learning.
  - Learning-Locality + Manifold Criterion.
  - Can handle continuous labels for supervision.
  - Discriminating power can be improved.
- DSA, CEA, and CTA are all supervised learning.
  - Learning-Locality + Multiple Sample Metric + Fisher Graph.
  - Learning-Locality + Fisher Graph + Correlation Metric.
  - Learning-Locality + Fisher Graph + Correlation Metric + High-order Data.
  - Can handle discrete labels for supervision.
  - Discriminating power is high.
- Computational Issue.
  - May suffer from computational difficulty when the sample space is large.
  - Using sample-locality to deal with the problem.
  - CEA and CTA may introduce more computational cost in the iteration process.
  - CTA may also reduce its computational costs owing to the reduced data dimensions in generalized eigen-decomposition.

Discussion 2

Table 8.1: Face recognition performance comparison on the PIE database.

Method	5 Train. Sample		10 Train. Sample		15 Train. Sample		20 Train. Sample	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
N-LEA-C	57.91%	142	26.35%	214	17.03%	209	12.82%	153
N-DSA-C	**34.03%**	169	20.65%	144	13.85%	177	11.13%	212
CEA	35.29%	299	**18.87%**	196	**13.62%**	81	**9.45%**	242
CTA	41.78%	16,16	24.39%	16,16	17.26%	14,14	12.82%	14,14

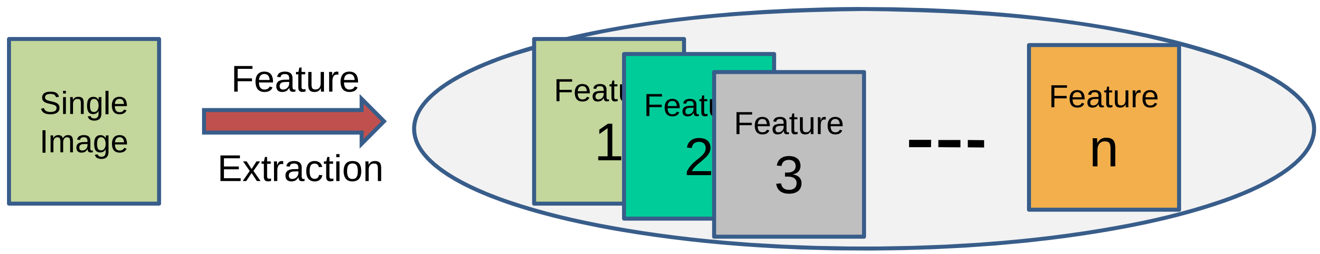
Table 8.2: Face recognition performance comparison on the Yale-B database.

Method	5 Train. Sample		10 Train. Sample		20 Train. Sample		30 Train. Sample	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
N-LEA-C	20.25%	109	8.87%	156	7.89%	271	8.82%	242
N-DSA-C	20.38%	84	**7.21%**	141	5.74%	166	5.80%	171
CEA	22.61%	103	7.55%	229	6.22%	218	3.56%	202
CTA	**16.99%**	26,26	7.60%	23,23	**4.96%**	23,23	**2.94%**	23,23

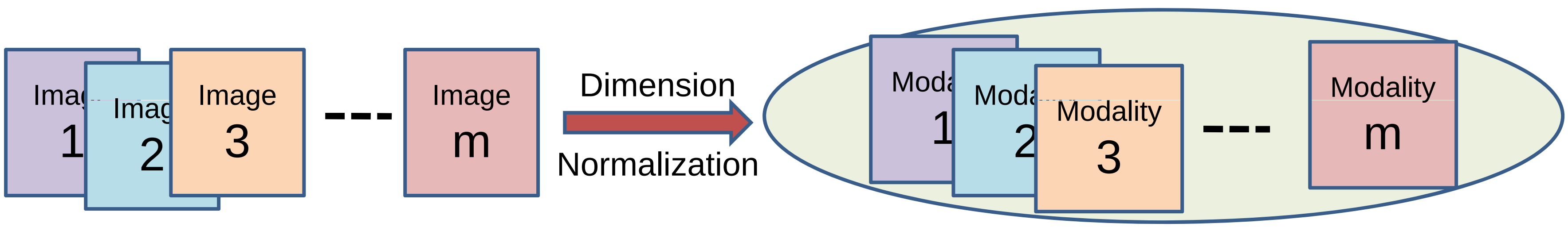
- CEA, CTA, and N-DSA-C all outperform N-LEA-C in all the cases.
- On the PIE database, CEA is more powerful than the other methods.
- On the PIE database, Yale-B database CTA performs the best.
- N-DSA-C may outperform the other methods when training sample size is small.

Pattern Fusion: Framework Extension

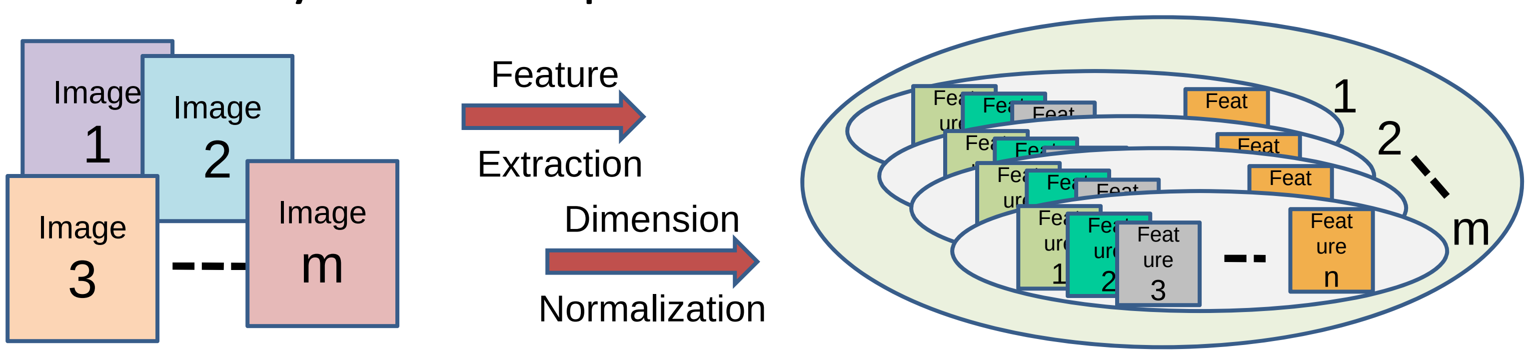
- Multiple Features



- Multimodality

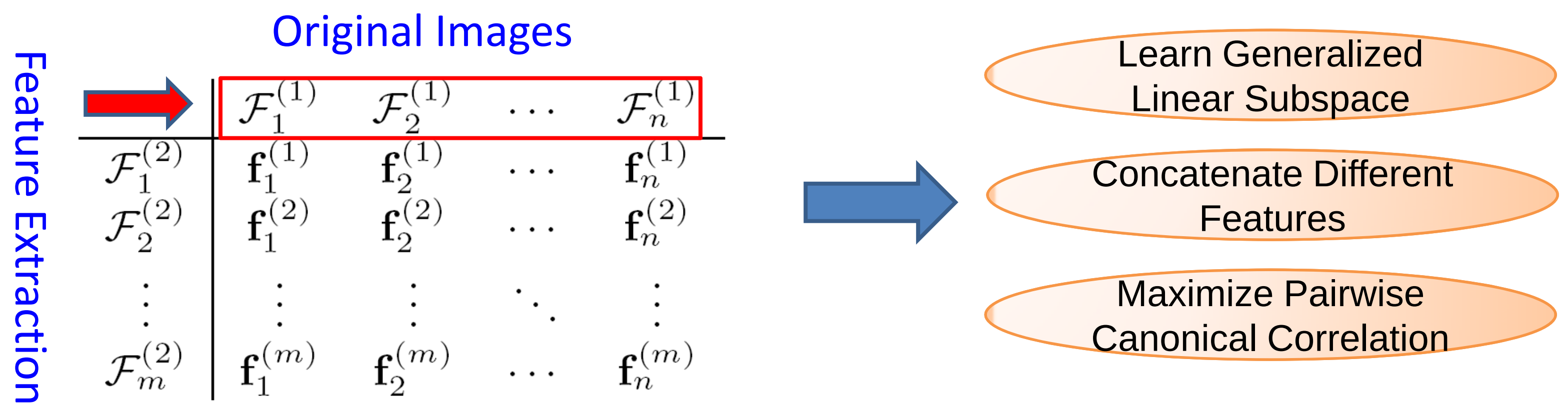


- Multimodality and Multiple Features



Multimodality Image Analysis

- Multiple Feature Fusion by F-Matrix



- Feature-Locality + Fisher Graph + Correlation Metric + High-order Data
- Feature Dimension Normalization
- Similarity Measure of Feature Sets [Kim, et. al, TPAMI2007]
  - Define feature matrices $\mathbf{F}_1$ and $\mathbf{F}_2$
  - Find $\mathbf{F}_1\mathbf{F}_1^T = \mathbf{P}_1\mathbf{\Lambda}_1\mathbf{P}_1^T$ and $\mathbf{F}_2\mathbf{F}_2^T = \mathbf{P}_2\mathbf{\Lambda}_2\mathbf{P}_2^T$
  - The SVD of $\mathbf{P}_1^T\mathbf{P}_2$ is $\mathbf{Q}_{12}\mathbf{\Lambda}_0\mathbf{Q}_{21}^T$
  - Canonical correlation similarity measure

$$S(\mathcal{F}_1, \mathcal{F}_2) = \max \ \mathrm{Tr}(\mathbf{Q}_{12}^T\mathbf{P}_1^T\mathbf{P}_2\mathbf{Q}_{21})$$

Y. Fu, et. al., ACM CIVR, 2008.

Multiple Feature Fusion in Subspace

- Unsupervised Multiple Feature Fusion

$$\mathbf{P} = \arg\max_{\mathbf{P}} \mathcal{J}_1 = \arg\max_{\mathbf{P}} \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} S_{\mathbf{P}}(\mathcal{F}_{k_1}^{(2)}, \mathcal{F}_{k_2}^{(2)}),$$

where

$$S_{\mathbf{P}}(\mathcal{F}_{k_1}^{(2)}, \mathcal{F}_{k_2}^{(2)}) = \max\ \mathrm{Tr}(\mathbf{Q}_{k_1 k_2}^T \mathbf{P}_{k_1}^T \mathbf{P}\mathbf{P}^T \mathbf{P}_{k_2} \mathbf{Q}_{k_2 k_1}).$$

- Supervised Multiple Feature Fusion

$$\begin{aligned} \mathbf{P} &= \arg\max_{\mathbf{P}} (\mathcal{J}_1 + \mathcal{J}_2) \\ &= \arg\max_{\mathbf{P}} (\mathcal{J}_1 + \sum_{i_1=1}^{n} \sum_{l_{i_2}=l_{i_1}} S_{\mathbf{P}}(\mathcal{F}_{i_1}^{(1)}, \mathcal{F}_{i_2}^{(1)}, \mathcal{L})), \end{aligned}$$

where

$$S_{\mathbf{P}}(\mathcal{F}_{i_1}^{(1)}, \mathcal{F}_{i_2}^{(1)}, \mathcal{L}) = \max\ \mathrm{Tr}(\mathbf{Q}_{i_1 i_2}^T \mathbf{P}_{i_1}^T \mathbf{P}\mathbf{P}^T \mathbf{P}_{i_2} \mathbf{Q}_{i_2 i_1}).$$

Y. Fu, et. al., ACM CIVR, 2008.

Multiple Feature Face Recognition

Table 1: Description of some acronyms and abbreviations.

Method	Description
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
Tensor	Tensor-based subspace learning
UnSL	Unsupervised Subspace Learning
SuSL	Supervised Subspace Leraning
Raw	Raw image feature
HoG	Histogram of Oriented Gradient feature
LBP	Local Binary Pattern feature
Fusion	Fusion of the three features
EuNN	Euclidean-distance Nearest Neighbor
CorrNN	Correlation-distance Nearest Neighbor

- *FRGC Ver1.0*: 275 subjects. 10 for training, 10 for test.
- *CMU PIE database*: 68 subjects. 20 for training, 20 for test.
- *Yale-B database*: 38 subjects. 20 for training, 20 for test.

Table 2: Single feature vs. multiple feature fusion.

	FRGC Ver1.0		PIE		Yale-B	
Feature	Accuracy	Dim.	Accuracy	Dim.	Accuracy	Dim.
Raw+EuNN	73.9%	240	58.8%	400	70.3%	400
HoG+EuNN	76.4%	380	63.7%	380	74.6%	500
LBP+EuNN	76.1%	500	72.1%	500	53.2%	400
Fusion+EuNN	79.4%	440	69.6%	470	71.1%	700
Raw+CorrNN	73.9%	280	58.8%	400	70.4%	480
HoG+CorrNN	76.4%	380	63.7%	380	75.0%	480
LBP+CorrNN	76.1%	500	72.0%	380	53.2%	300
Fusion+CorrNN	79.3%	380	69.6%	500	70.9%	540

Y. Fu, et. al., ACM CIVR, 2008.

Multiple Feature Face Recognition

$$
\begin{array}{c|cccc}
 & \mathcal{F}_1^{(1)} & \mathcal{F}_2^{(1)} & \cdots & \mathcal{F}_n^{(1)} \\
\hline
\mathcal{F}_1^{(2)} & \mathbf{f}_1^{(1)} & \mathbf{f}_2^{(1)} & \cdots & \mathbf{f}_n^{(1)} \\
\mathcal{F}_2^{(2)} & \mathbf{f}_1^{(2)} & \mathbf{f}_2^{(2)} & \cdots & \mathbf{f}_n^{(2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathcal{F}_m^{(2)} & \mathbf{f}_1^{(m)} & \mathbf{f}_2^{(m)} & \cdots & \mathbf{f}_n^{(m)}
\end{array}
$$

Use CTA for further feature extraction.

Table 1: Description of some acronyms and abbreviations.

Method	Description
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
Tensor	Tensor-based subspace learning
UnSL	Unsupervised Subspace Learning
SuSL	Supervised Subspace Leraning
Raw	Raw image feature
HoG	Histogram of Oriented Gradient feature
LBP	Local Binary Pattern feature
Fusion	Fusion of the three features
EuNN	Euclidean-distance Nearest Neighbor
CorrNN	Correlation-distance Nearest Neighbor

Table 3: Multiple feature fusion by subspace learning.

	PIE		Yale-B	
Method	Accuracy	Dim.	Accuracy	Dim.
PCA+EuNN	69.6%	470	71.1%	700
PCA+CorrNN	69.6%	500	70.9%	540
UnSL+EuNN	71.2%	340	71.8%	310
UnSL+CorrNN	71.3%	350	71.6%	310
SuSL+EuNN	71.5%	400	85.4%	300
SuSL+CorrNN	93.6%	390	94.3%	390

Table 4: Combine multiple feature fusion with tensor-based discriminant analysis.

	PIE		Yale-B	
Method	Accuracy	Dim.	Accuracy	Dim.
PCA+LDA	92.8%	65	94.1%	35
UnSL+LDA	93.0%	65	96.6%	30
SuSL+LDA	92.8%	65	97.2%	35
UnSL+Tensor	94.7%	70×3	97.4%	34×3
SuSL+Tensor	94.3%	70×3	97.5%	37×3

Y. Fu, et. al., ACM CIVR, 2008.

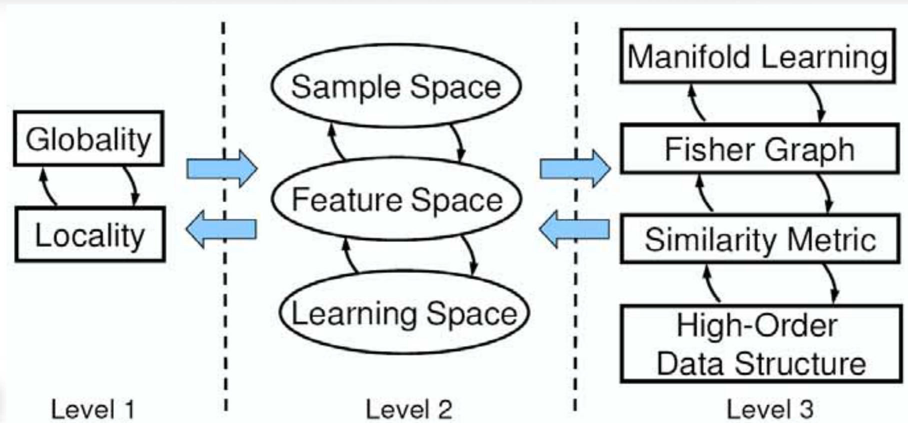
Discussion 3

- Simply concatenating different features may improve the robustness of face recognition performance.
- The feature fusion may also degrade the performance due to the unbalance among the individual features.
- The proposed method learns a generalized subspace in which the low-dimensional representations of those individual features have a better balance to contribute to the improved performance by fusion.
- CTA is applied for following reason.
  - Capture the high-order feature patterns for fusion.
  - Reduce the computational cost when the number of different features is large.
  - Alleviate the curse-of-dimensionality dilemma and the small sample size problem.
- A non-linear learning strategy is also feasible to extend if we assume the correlations among different features tend to be more complicated.

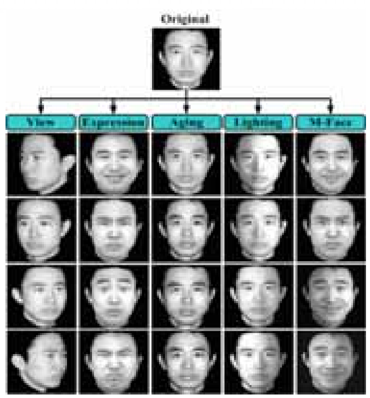
Summary

Theoretical-Driven Research

Machine Learning



Facial Image Computing



Reverse

Face Recognition

Facial Expression Analysis

Face Pose Estimation

Age Estimation

Lipreading

Application-Driven Research

Future Work

- Multi-Scale and Multimodality Biomedical Image Fusion
  - Beckman Graduate Fellowship (PI)
  - 3D breast model and multimodality data generation
  - Breast cancer detection with multimodality image fusion
  - Collaborated with Prof. Michael Insana (Ultrasound), Prof. Zhi-pei Liang (MRI), and Prof. Rohit Bhargava (FTIR)
- Machine Learning and Pattern Recognition
  - Biometrics
  - Multi-task multi-instance multi-label learning
  - Semi-supervised metric learning and ranking
  - Sparse representation, non-negative graph embedding
  - Classifier ensembles
  - Computational biology and bioinformatics: Mining various sources of high-throughput data (e.g. protein-protein interaction data, microarray gene expression data, genome sequences, etc.)

Future Work (cont.)

- Interactive Multimedia Systems
  - Human-in-the-loop (e.g. semi-supervised clustering/labeling)
  - Human-centered content and context modeling for multimedia information retrieval
  - Integration of Context and Content for Multimedia Management
  - Web-context for online multimedia annotation, browsing, sharing and reuse
  - Avatar-based communication systems
- Computer Vision and Image Processing
  - Automatic alignment
  - Detection/tracking
  - Spatio-temporal analysis
  - Image-based general object classification
  - Event representation and motion trajectory analysis
  - Audio-visual intelligent system (e.g. lipreading, person identification)

Collaborators

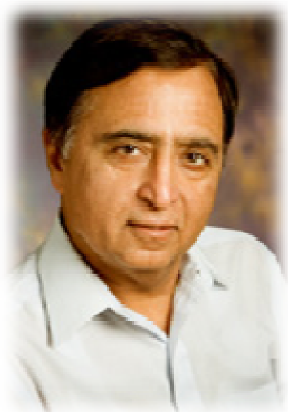
Prof. Thomas Huang
UIUC

Prof. Robert Fossum
UIUC

Prof. Michael Insana
UIUC

Prof. Mark Hasegawa-Johnson, UIUC

Prof. Nanning Zheng
XJTU, China



Prof. Narendra Ahuja
UIUC

Prof. Charles Dyer
Uni. of Wisconsin

Prof. Aggelos Katsaggelos
Northwestern University

Collaborators (cont.)

Academia

- Prof. Ying Wu, EECS, Northwestern University
- Prof. Rohit Bhargava, Bioengineering, UIUC
- Prof. Zhi-pei Liang, ECE, UIUC
- Prof. Michelle Wang, Statistics, Psychology, and Bioengineering, UIUC
- Prof. Qi Tian, Computer Science, University of Texas at San Antonio
- Prof. Guodong Guo, Mathematics and CS, North Carolina Central University
- Prof. Tony X. Han, ECE, University of Missouri-Columbia
- Prof. Xuelong Li, Computer Science, University of London
- Prof. James Zhu Li, Dept. of Computing, Hong Kong Polytechnic University
- Prof. Shuicheng Yan, ECE, National University of Singapore
- Dr. Zhihong Zeng, Beckman Postdoctoral Fellow, UIUC
- Dr. Ranga C. Gudivada, Biomedical Informatics, University of Cincinnati

Industry

- Dr. Ming Liu, Researcher, Microsoft Research
- Dr. Richard Li, Principal Staff Research Engineer, Motorola Research Lab
- Dr. Jilin Tu, Computer Scientist, GE Global Research
- Dr. Baback Moghaddam, Principal Member, Jet Propulsion Laboratory
- Dr. Wei Qu, Senior Researcher, Siemens Medical Solutions USA Inc.
- Dr. Derek Li, Chairman and Founder, Zienon LLC
- Dr. Zhen Wen, Research Staff Member, IBM T.J.Watson Research Center
- Dr. Jerry Yu, Researcher, Kodak Research Labs
- Dr. Shyamsundar Rajaram, Research Scientist, Hewlett-Packard Labs

Demos

- M-Face and FaceTransfer

  http://www.ifp.uiuc.edu/~yunfu2/M-Face.html

  http://www.ifp.uiuc.edu/~yunfu2/FaceTransfer.html
- RTM-HAI: Real-Time Multimodal Human-Avatar Interaction
  http://www.ifp.uiuc.edu/~yunfu2/RTM-HAI.html
- hMouse

  http://www.ifp.uiuc.edu/~yunfu2/
- Realtime Shrug Detector

  http://www.ifp.uiuc.edu/~hning2/shrug.htm
- EAVA: A 3D Emotive Audio-Visual Avatar

  http://www.ifp.uiuc.edu/~haotang2/peava.html