ECE Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN | COLLEGE OF ENGINEERING

Non-Negative Graph Embedding

Jianchao Yang, Shuicheng Yan, Yun Fu,
Xuelong Li, Thomas Huang

Department of ECE, Beckman Institute and CSL

University of Illinois at Urbana-Champaign

COORDINATED SCIENCE LABORATORY

BECKMAN INSTITUTE AT THE UNIVERSITY OF ILLINOIS

Outline

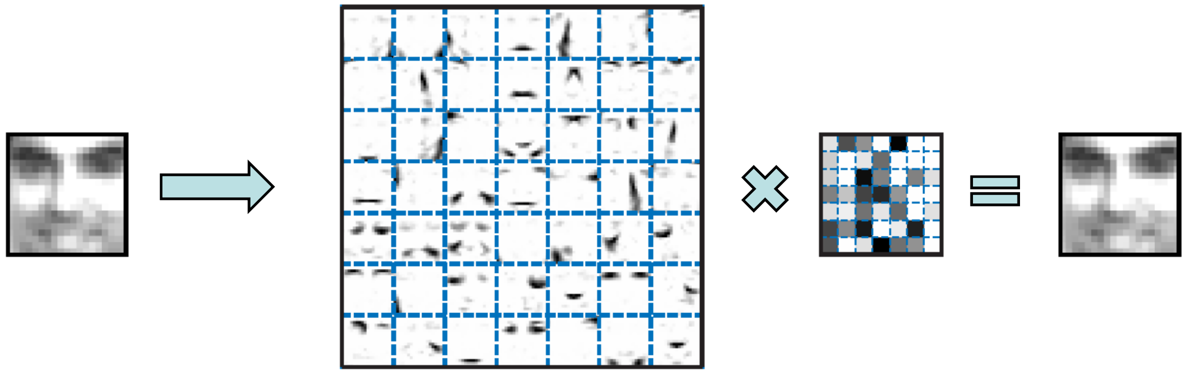
- Non-negative Part-based Representation
  - Non-Negative Matrix Factorization
- Non-negative Graph Embedding (NGE)
  - Graph Embedding framework
  - Our formulation
- Experiment Results
  - Face recognition
  - Localized basis
  - Robust to image occlusion
- Conclusions

Non-negative Part-based Representation

$$Object_i = b_{i1} \times Part_1 + b_{i2} \times Part_2 + \ldots$$

- Why non-negativity?
  - Better physical interpretation of the non-negative data
  - Examples such as absolute temperatures, light intensities, probabilities, sound spectra, etc.
- Why part-based?
  - Psychological and physiological evidence for part-based representations in the human brain.
  - Perception of the whole as perceptions of the parts.

Non-negative Matrix Factorization



- Formulation

$$\arg\min_{W,H} \|D - WH\|_2^2 \quad s.t. \quad W \geq 0, H \geq 0$$

- Multiplicative update rules guarantee non-negativity

$$H_{ij} \longleftarrow H_{ij} \frac{(W^T D)_{ij}}{(W^T W H)_{ij}} \qquad W_{ti} \longleftarrow W_{ti} \frac{(D H^T)_{ti}}{(W H H^T)_{ki}}$$

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008

What NMF Learns?

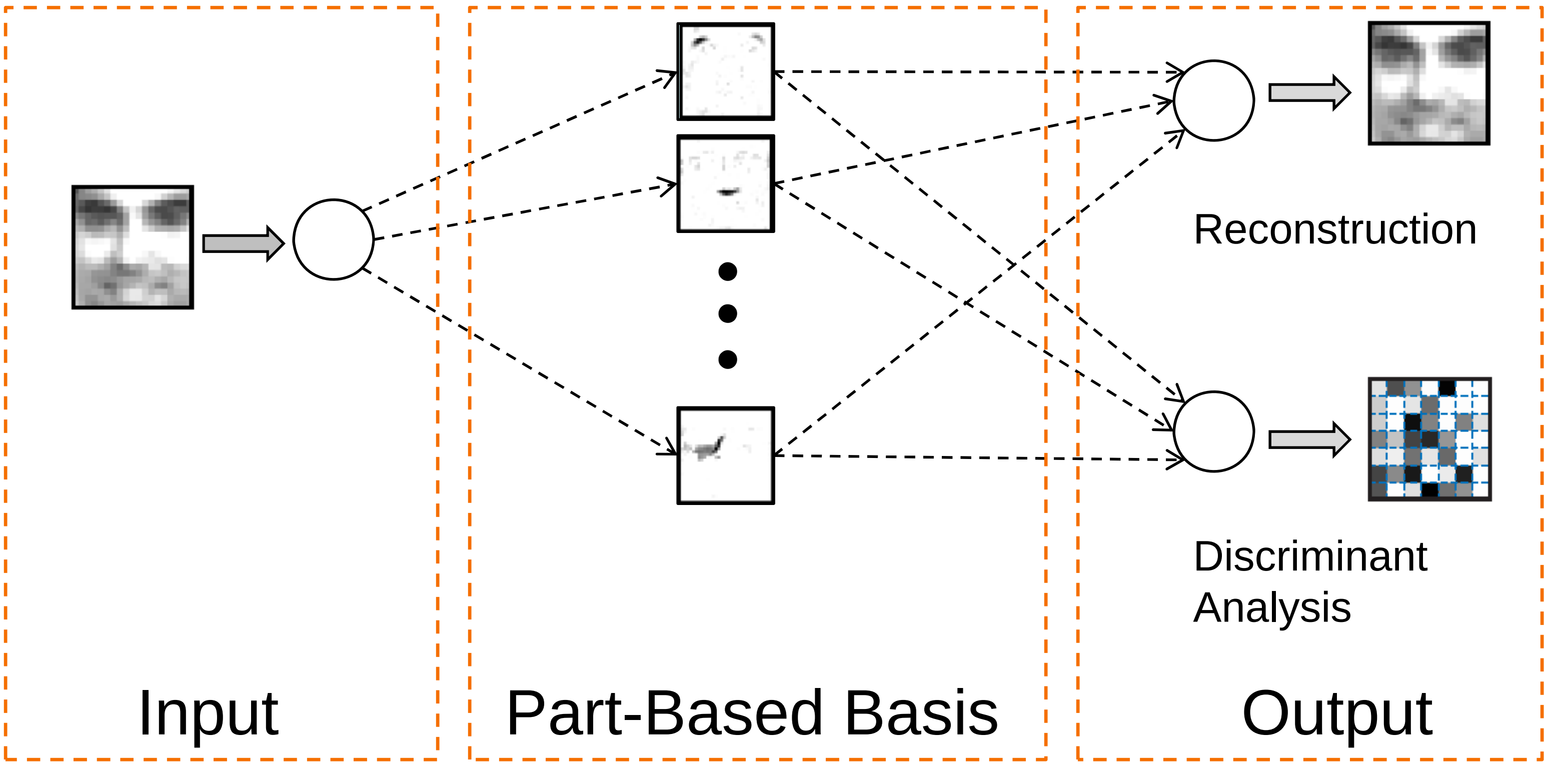
- NMF indeed learns part-based representation.
- Problems:
  - Matrix factorization has no control on the properties of the parts.
  - Used in document clustering, but not good for recognition.
- How the brain learns the discriminative parts is still unknown.

Non-negative Graph Embedding (NGE)

- Motivation
  - Learn the non-negative part-based representation
  - Want it to be good for classification

- Method
  - Reconstruction for learning the part-based basis
  - Regularization with discriminant analysis

A Better Scheme



Use all available data for learning the basis, while guided by the labeling information.

Learn the Discriminative Parts

- One straightforward solution

$$\min_{W,H} \|X - WH\|_2^2 + \lambda f(H), \quad \text{s.t.} \quad W, H \geq 0$$

X : the data matrix

W : the part-based basis matrix

H : the coefficient matrix

$f(H)$: function encoding the discriminative power of coefficients

- The problem is how to choose $f(H)$ and to do the optimization.

Graph Embedding

- Graph Embedding Framework [Yan, et al 2007]
  - Intrinsic Graph: characterize the favorable relationship among training data.
    $$G = \{X, S\}$$
  - Penalty Graph: characterize the unfavorable relationship among training data
    $$G^p = \{X, S^p\}$$
  - Objective:
    $$\begin{cases} \max_H \sum_{i \neq j} \|h_i - h_j\|^2 S_{ij}^p \\ \min_H \sum_{i \neq j} \|h_i - h_j\|^2 S_{ij} \end{cases}$$
- These graphs can be unsupervised, supervised or semi-supervised.

NGE Formulation

- Divide the feature space into two parts--discriminant space H^1 and the complementary space H^2 for reconstruction.

$$H = \begin{bmatrix} H^1 \\ H^2 \end{bmatrix} \qquad W = [W^1 \; W^2]$$

- The objective for H^1 is:

$$\begin{cases} \max_{H^1} \sum_{i \neq j} \|h_i^1 - h_j^1\|^2 S_{ij}^p \\ \min_{H^1} \sum_{i \neq j} \|h_i^1 - h_j^1\|^2 S_{ij} \end{cases}$$

NGE Formulation

- To make the problem solvable, change the objective with the complementary space:

$$\begin{cases} \min_{H^2} \sum_{i \neq j} \|h_i^2 - h_j^2\|^2 S_{ij}^p \\ \min_{H^1} \sum_{i \neq j} \|h_i^1 - h_j^1\|^2 S_{ij} \end{cases}$$

- Given the intrinsic graph and penalty graph, the optimization problem can formulated as:

$$\min_{W,H} \sum_{i \neq j} \|h_i^1 - h_j^1\|^2 S_{ij} + \sum_{i \neq j} \|h_i^2 - h_j^2\|^2 S_{ij}^p$$
$$+\lambda \|X - WH\|^2, \;\; s.t. \;\; W, H \geq 0$$

Preliminaries

- **Definition 1:** A matrix ***B*** is called *M*-matrix if 1) the off-diagonal entries are less than or equal to zeros; 2) the real parts of all eigen values are positive.
- **Lemma 1:** If ***B*** is a *M*-matrix, its inverse is non-negative, that is ***B***(i,j) >= 0.

- **Definition 2:** Function G(A, A') is an auxiliary function for F(A) if G(A, A')>= F(A) and G(A, A) = F(A).
- **Lemma 2:** If G is an auxiliary function of F, F is non-increasing under the following update rule:

$$A^{t+1} = \arg\min_{A} G(A, A^{t})$$

Optimization Procedure

- **Initialize** W and H with non-negative values, and the optimization is done by alternating between W and H.

- **Optimize W, fixing H**. Define the auxiliary function as

$$G(W_i, W_i^t) = W_i D^h W_i^T + f(W_i^t) + \nabla f(W_i^t)(W_i - W_i^t)^T + \frac{1}{2}(W_i - W_i^t) K(W_i^t)(W_i - W_i^t)^T,$$

Thus the update rule for W is:

$$W_i^{t+1} = \lambda X_i H^T (K(W_i^t) + 2D^h)^{-1}$$

where $K(W_i^t) + 2D^h$ is a diagonal element-wise positive matrix, which guarantees the non-negativity of W.

Optimization Procedure

- **Optimize H, fixing W.** The auxiliary function is defined as

$$G(H, H^t) = Tr(H^1 L H^{1^T}) + Tr(H^2 L^p H^{2^T}) + g(H, H^t)$$

  – To optimize H^1:

$$H_i = \lambda\, w_i^T\, X (K^i + 2L)^{-1}$$

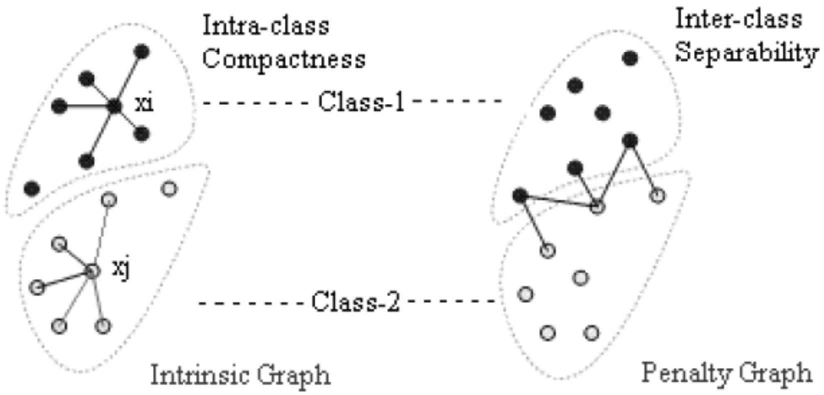
  – To optimize H^2:

$$H_i = \lambda\, w_i^T\, X (K^i + 2L^p)^{-1}$$

$K^i + 2L$ and $K^i + 2L^p$ are *M*-matrix, whose inverse are element-wise non-negative, hence guarantees non-negativity of H.

General Framework

Intrinsic and penalty graphs for Marginal Fisher Analysis



- Our algorithm is a general framework, given the intrinsic and penalty graphs.
  - These graphs can be unsupervised, supervised or semi-supervised.
  - We used **supervised** Marginal Fisher Analysis (MFA) **graph** to demonstrate the framework.

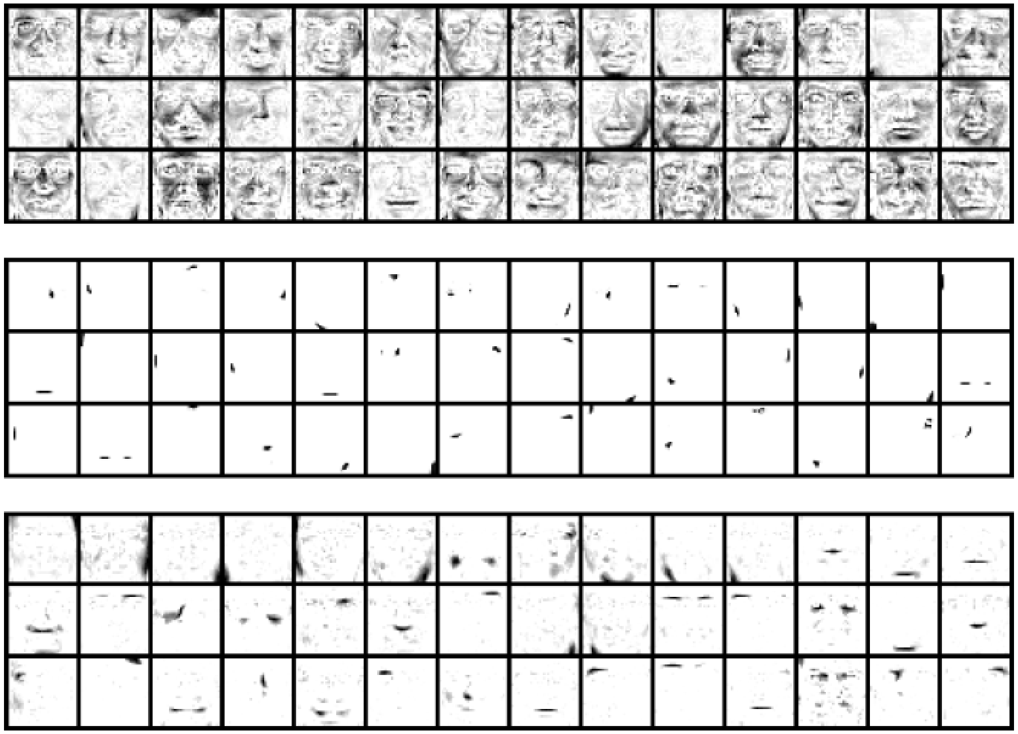
Experiments

- Face Recognition
  - Tested on three databases: CMU PIE, ORL and FERET.
  - Compared with unsupervised algorithms PCA, NMF, LNMF (S. Li, CVPR 2001) and supervised algorithms LDA and MFA.

Algorithm	PIE	ORL	FERET
Baseline	82.54	85.00	81.90
PCA	82.54 (124)	85.50 (105)	81.90 (141)
NMF	80.67 (208)	74.00 (158)	83.81 (174)
LNMF	92.06 (108)	87.50 (130)	81.90 (172)
NGE	**98.10 (127)**	**95.50 (121)**	**92.42 (152)**
LDA	94.92 (62)	94.50 (39)	91.60 (67)
MFA	95.87 (116)	95.50 (48)	91.43 (43)

Experiments

- Learned non-negative part-based basis



NMF

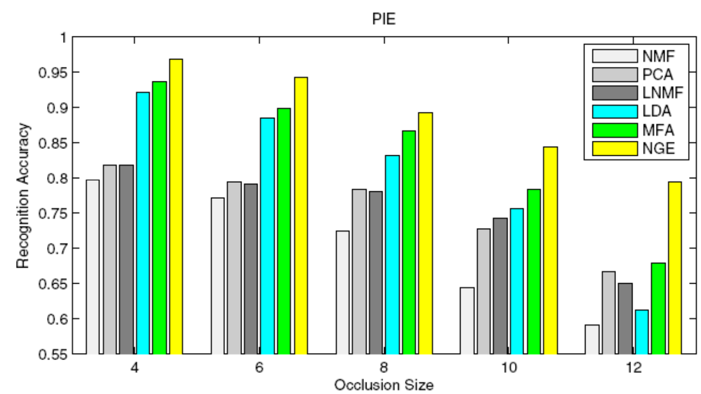
LNMF

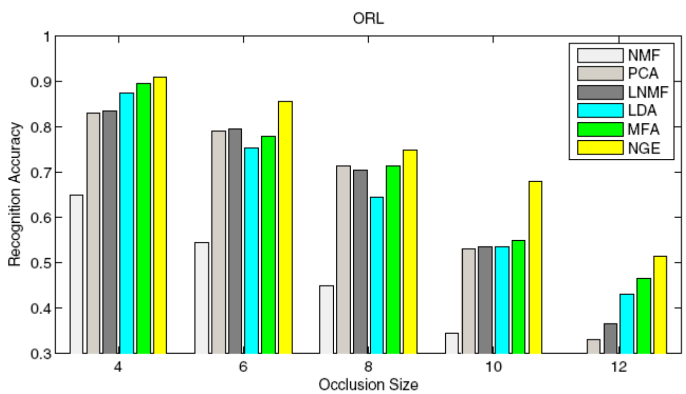
NGE

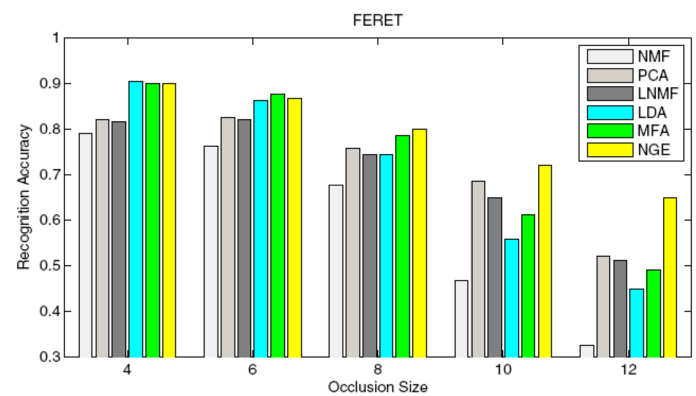
Experiments

- Robust to Occlusion

Occlusion Examples







IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008

Conclusions

- Contributions:
  - Proposed a general framework called Non-Negative Graph Embedding (NGE).
  - **Supervised** MFA **graph** is used to demonstrate the effectiveness of the algorithm.
- Limitation:
  - Like other graph-based method, NGE suffers from speed and scalability during the off-line training.
- Extension:
  - Unlabeled data can be incorporated into the basis learning, while guided by the available label information.

Thank you!