Supplementary Material:
Projective Unsupervised Flexible Embedding with Optimal Graph

7 Benchmark Dataset Details

To evaluate our model, we choose three multiview action recognition datasets, three face datasets and one handwritten digit recognition dataset.

The **action** datasets are the IXMAS dataset [27], the newer version of of IXMAS dataset referred to as NIXMAS and and the partially occluded dataset OIXMAS [28] dataset. IXMAS dataset consists of 12 action classes. Each action is performed 3 times by 11 actors and recorded from 5 different viewpoints. The OIXMAS dataset has the same set of actions as IXMAS. But parts of the actions are occluded. The self-similarity matrix features are extracted for each video.

The three **face** datasets are the JAFFE dataset, the UMIST face dataset, and the YaleB dataset. The JAFFE dataset [13] contains 213 images of different facial expressions from 10 different Japanese female models. The images are resized to 26×26 and represented by pixel values. The UMIST face dataset [4] contains 564 images of 20 individuals. Each individual has photos in different poses. The concatenated pixel values are used as features. The YaleB dataset [3] contains 2414 near frontal images from 38 persons under different illuminations. Each image is resized to 32×32 and the pixel value is used as feature representation.

We also use the USPS [9] dataset to validate the performance on **handwritten digit** recognition. The dataset consists of 9298 gray-scale handwritten digit images. We resize the images to 16×16 and employ the pixel values as features.

8 Deduction of Eq. (2)

The deduction of the manifold term reformulation is as follows:

$$\begin{aligned}
&\sum_{i,j=1}^{n} A_{i,j}\|\boldsymbol{f}_i-\boldsymbol{f}_j\|_2^2 = \sum_{i,j=1}^{n} A_{i,j}(\boldsymbol{f}_i^2+\boldsymbol{f}_j^2-2\boldsymbol{f}_i\boldsymbol{f}_j)\\
=&\sum_i \boldsymbol{f}_i^2\boldsymbol{D}_{i,i}+\sum_j \boldsymbol{f}_j^2\boldsymbol{D}_{j,j}-2\sum_{i,j}\boldsymbol{f}_i\boldsymbol{f}_j A_{i,j}\\
=&2Tr(\boldsymbol{F}^T\boldsymbol{L}_A\boldsymbol{F})
\end{aligned}\tag{15}$$

9 Deduction of Eq. (10)

For any d, we define $\mathbf{1}_d \in \mathbb{R}^d$ as a column vector with all its elements equal to 1. Then $\boldsymbol{H}_n=\boldsymbol{I}-(1/n)\mathbf{1}_n\mathbf{1}_n^T \in \mathbb{R}^{n\times n}$ is a matrix for centering the data by subtracting the mean of the data. Put $\boldsymbol{W},\boldsymbol{b}$ into the regression function, we have

$$\begin{aligned}
\boldsymbol{X}^T\boldsymbol{W}+\mathbf{1}\boldsymbol{b}^T=&\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{F}+\frac{1}{n}\mathbf{1}\mathbf{1}^T\boldsymbol{F}-\frac{1}{n}\mathbf{1}\mathbf{1}^T\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{F}\\
=&\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{F}+\frac{1}{n}\mathbf{1}\mathbf{1}^T\boldsymbol{F}=\boldsymbol{B}\boldsymbol{F}
\end{aligned}\tag{16}$$

where $\boldsymbol{B}{=}\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}+\frac{1}{n}\mathbf{1}\mathbf{1}^T$. Put $\boldsymbol{W}$,$\boldsymbol{b}$ to Eq. (8), we have

$$\boldsymbol{F}^*{=}arg\min_{\substack{\boldsymbol{F}\in\mathbb{R}^{n\times c}\\ \boldsymbol{F}^T\boldsymbol{F}=\boldsymbol{I}}}\begin{pmatrix}\lambda Tr(\boldsymbol{F}^T\boldsymbol{L_S}\boldsymbol{F}){+}\gamma Tr(\boldsymbol{F}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{F}),\\ +\mu Tr(\boldsymbol{B}\boldsymbol{F}-\boldsymbol{F})^T(\boldsymbol{B}\boldsymbol{F}-\boldsymbol{F})\end{pmatrix} \tag{17}$$

The term in Eq. (17) can be reformulated as

$$\begin{aligned}&Tr(\boldsymbol{F}^T(\lambda\boldsymbol{L_S})\boldsymbol{F}){+}Tr(\boldsymbol{F}^T(\gamma\boldsymbol{Q}^T\boldsymbol{Q})\boldsymbol{F})\\ &+Tr(\boldsymbol{F}^T(\mu(\boldsymbol{B}-\boldsymbol{I})^T(\boldsymbol{B}-\boldsymbol{I}))\boldsymbol{F})\\ {=}&Tr(\boldsymbol{F}^T(\lambda\boldsymbol{L_S}{+}\gamma\boldsymbol{Q}^T\boldsymbol{Q}+\mu(\boldsymbol{B}-\boldsymbol{I})^T(\boldsymbol{B}-\boldsymbol{I}))\boldsymbol{F})\end{aligned} \tag{18}$$

Using $\boldsymbol{H}_n\boldsymbol{H}_n=\boldsymbol{H}_n=\boldsymbol{H}_n^T$, and $\mu\boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}+\gamma\boldsymbol{Q}^T\boldsymbol{Q}=\mu\boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{H}_n$. The term

$$\gamma\boldsymbol{Q}^T\boldsymbol{Q}+\mu(\boldsymbol{B}-\boldsymbol{I})^T(\boldsymbol{B}-\boldsymbol{I})$$

in Eq. (18) can be written as

$$\begin{aligned}&\gamma\boldsymbol{Q}^T\boldsymbol{Q}+\mu(\boldsymbol{Q}^T\boldsymbol{X}-\boldsymbol{I})\boldsymbol{H}_n(\boldsymbol{X}^T\boldsymbol{Q}-\boldsymbol{I})\\ {=}&\gamma\boldsymbol{Q}^T\boldsymbol{Q}{+}\mu(\boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}{-}2\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}{+}\boldsymbol{H}_n)\\ {=}&\mu\boldsymbol{H}_n-\mu\boldsymbol{H}_n\boldsymbol{X}^T\boldsymbol{Q}\\ {=}&\mu\boldsymbol{H}_n-\mu\boldsymbol{H}_n\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{H}_n\boldsymbol{X}^T+\frac{\gamma}{\mu}\boldsymbol{I})^{-1}\boldsymbol{X}\boldsymbol{H}_n\end{aligned} \tag{19}$$

Let $\boldsymbol{X}_n=\boldsymbol{X}\boldsymbol{H}_n$. According to Eq. (19), we rewrite Eq. (17) as

$$\boldsymbol{F}^*{=}arg\min_{\boldsymbol{F},\boldsymbol{F}^T\boldsymbol{F}=\boldsymbol{I}} Tr\boldsymbol{F}^T(\lambda\boldsymbol{L_S}+\mu\boldsymbol{H}_n-\mu\boldsymbol{N})\boldsymbol{F} \tag{20}$$

where $\boldsymbol{N}{=}\boldsymbol{X}_n^T(\boldsymbol{X}_n\boldsymbol{X}_n^T+\frac{\gamma}{\mu}\boldsymbol{I})^{-1}\boldsymbol{X}_n{=}\boldsymbol{X}_n^T\boldsymbol{X}_n(\boldsymbol{X}_n^T\boldsymbol{X}_n{+}\frac{\gamma}{\mu}\boldsymbol{I})^{-1}$ A generalized eigenvalue decomposition [31] can be utilized to solve this objective function.