Low-rank Matrix Factorization under General Mixture Noise Distributions

Xiangyong Cao, Yang Chen, Qian Zhao, Deyu Meng,
Yao Wang, Dong Wang, Zongben Xu

Xi'an Jiaotong Univerisity

Speaker: Zhuwen Li

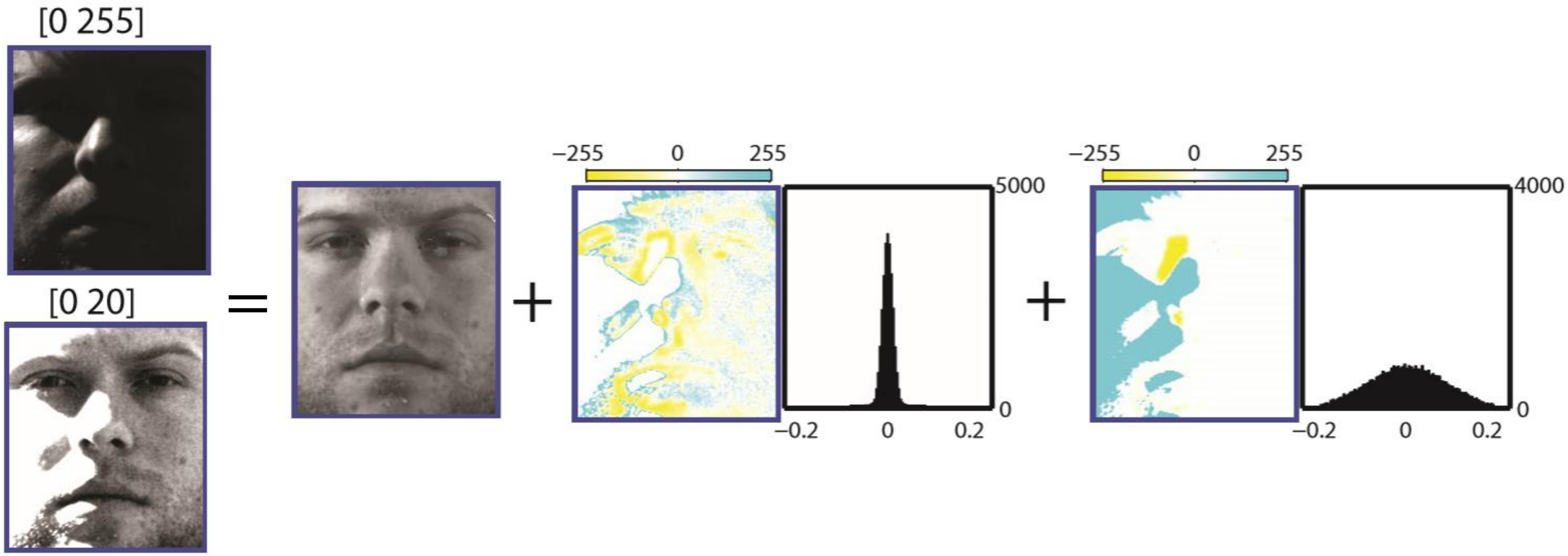
Introduction

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{W} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T)||_\ell$$

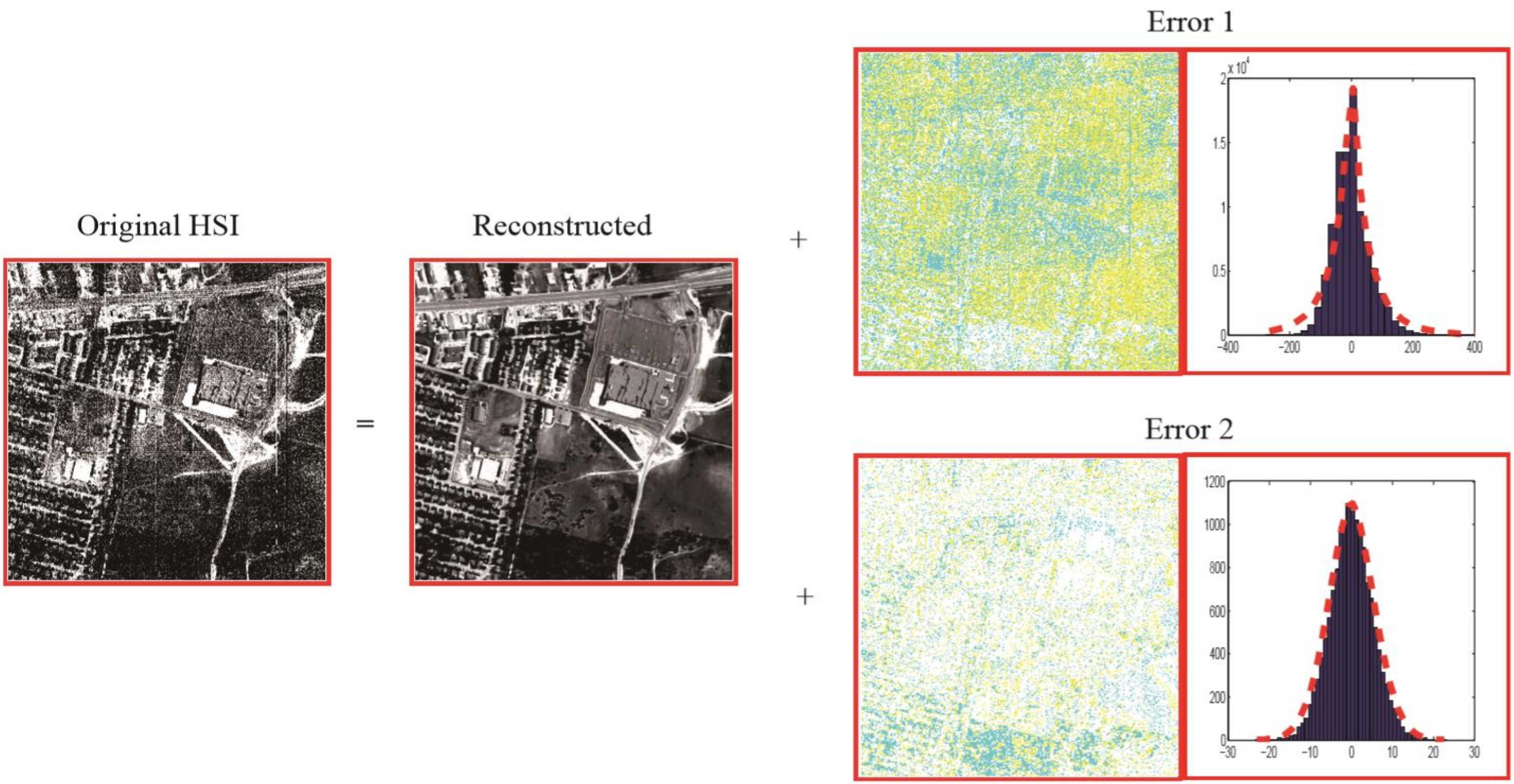
- $\ell = 1$, Laplacian Noise
- $\ell = 2$, Gaussian Noise

Mixture of Gaussian

- D. Meng and F. D. L. Torre. Robust matrix factorization with unknown noise. In ICCV. 2013.



Mixture of Exponential Power



Mixture of Exponential Power

$$y_{ij} = \mathbf{u}_i \mathbf{v}_j^T + e_{ij}$$

$$\mathbb{P}(e_{ij}) = \sum_{k=1}^{K} \pi_k f_{p_k}(e_{ij}; 0, \eta_k)$$

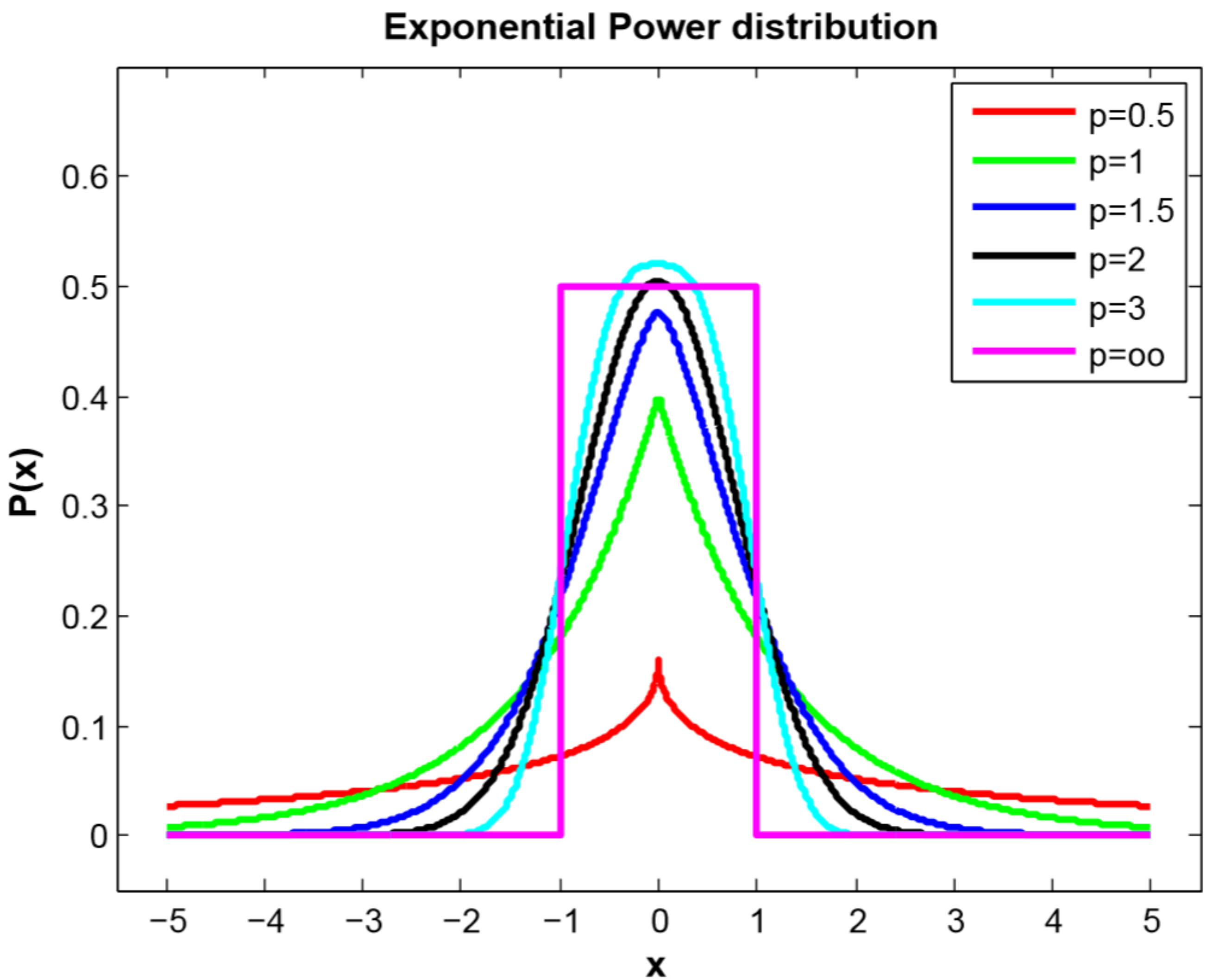
Mixture of Exponential Power

$$\mathbb{P}(e_{ij}) = \sum_{k=1}^{K} \pi_k f_{p_k}(e_{ij}; 0, \eta_k)$$

$$f_p(e; 0, \eta) \quad = \quad \frac{p\eta^{\frac{1}{p}}}{2\Gamma(\frac{1}{p})} \exp\{-\eta|e|^p\}$$

$$\eta = 1/(p\sigma^p)$$

Mixture of Exponential Power



Mixture of Exponential Power

- Indicator variable

$$\mathbf{z}_{ij} \;=\; [z_{ij1}, z_{ij2}, \ldots, z_{ijK}]^T$$

where $z_{ijk} \in \{0,1\}$ and $\sum_{k=1}^{K} z_{ijk} = 1.$

- $\mathbf{z}_{ij}$ obey a multinomial distribution

$$\mathbf{z}_{ij} \;\sim\; \mathcal{M}(\boldsymbol{\pi})$$

where $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_K]^T$

Mixture of Exponential Power

$$
\begin{aligned}
\mathbb{P}(e_{ij}|\mathbf{z}_{ij}) &= \prod_{k=1}^{K} f_{p_k}(e_{ij};0,\eta_k)^{z_{ijk}}, \\
\mathbb{P}(\mathbf{z}_{ij};\boldsymbol{\pi}) &= \prod_{k=1}^{K} \pi_k^{z_{ijk}}.
\end{aligned}
$$

Mixture of Exponential Power

- The likelihood function

$$\mathbb{P}(\mathbf{E}, \mathbf{Z}; \boldsymbol{\Theta}) \quad = \quad \prod_{i,j\in\Omega} \prod_{k=1}^{K} [\pi_k f_{p_k}(e_{ij}; 0, \eta_k)]^{z_{ijk}}$$

$$\mathbf{E} \quad = \quad (e_{ij})_{m\times n}$$

$$\mathbf{Z} \quad = \quad (\mathbf{z}_{ij})_{m\times n}$$

$$\boldsymbol{\Theta} \quad = \quad \{\boldsymbol{\pi}, \boldsymbol{\eta}, \mathbf{U}, \mathbf{V}\}$$

Mixture of Exponential Power

- The log-likelihood function

$$\begin{aligned} l^C(\mathbf{\Theta}) &= \log \mathbb{P}(\mathbf{E}, \mathbf{Z}; \mathbf{\Theta}) \\ &= \sum_{i,j \in \Omega} \sum_{k=1}^{K} z_{ijk} [\log \pi_k + \log f_{p_k}(e_{ij}; 0, \eta_k)] \end{aligned}$$

- Model selection (K)

$$\max_{\mathbf{\Theta}} \left\{ l_P^C(\mathbf{\Theta}) = l^C(\mathbf{\Theta}) - P(\boldsymbol{\pi}; \lambda) \right\}$$

$$P(\boldsymbol{\pi}; \lambda) = n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}$$

EM algorithm

- E-Step (conditional expectation of z_{ijk})

$$\gamma_{ijk}^{(t+1)} = \frac{\pi_k^{(t)} f_{p_k}(y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T)|0, \eta_k^{(t)})}{\sum_{l=1}^{K} \pi_l^{(t)} f_{p_l}(y_{ij} - \mathbf{u}_i^{(t)}(\mathbf{v}_j^{(t)})^T)|0, \eta_l^{(t)}))}$$

EM algorithm

- M-Step

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}) = \sum_{i,j \in \Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k]$$

$$- n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

EM algorithm

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k] - n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

- Update $\boldsymbol{\pi}$

$$\boldsymbol{L}(\boldsymbol{\pi}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} \log \pi_k - n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon} + \tau(\sum_k \pi_k - 1).$$

  - ➢ Take the derivative w.r.t $\boldsymbol{\pi}$ and set to zero

$$\pi_k^{(t+1)} = \max\left\{0, \frac{1}{1 - \lambda\hat{D}} \left[\frac{\sum_{i,j\in\Omega} \gamma_{ijk}^{(t+1)}}{|\Omega|} - \lambda D_k\right]\right\}$$

EM algorithm

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k]$$
$$- n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

- Update $\boldsymbol{\eta}$
  - ➢ Take the derivative w.r.t. η_k and set to zero

$$\eta_k^{(t+1)} = \frac{N_k}{p_k \sum_{i,j\in\Omega} \gamma_{ijk}^{(t+1)} |y_{ij} - \mathbf{u}_i^{(t)} (\mathbf{v}_j^{(t)})^T|^{p_k}}$$

EM algorithm

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k] - n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

- Update $\mathbf{U}, \mathbf{V}$

$$\sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} \log f_k(y_{ij} - \mathbf{u}_i^{(t)} (\mathbf{v}_j^{(t)})^T; \eta_k^{(t+1)})$$

$$\min_{\mathbf{U},\mathbf{V}} \sum_{k=1}^{K} ||\mathbf{W}_{(k)} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T)||_{p_k}^{p_k}$$

$$w_{(k)ij} = \begin{cases} (\eta_k^{(t+1)} \gamma_{ijk}^{(t+1)})^{\frac{1}{p_k}}, & i, j \in \Omega \\ 0, & i, j \notin \Omega \end{cases}$$

EM algorithm

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}) = \sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk}^{(t+1)} [\log f_{p_k}(e_{ij}; 0, \eta_k) + \log \pi_k] - n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.$$

- Update $\mathbf{U}, \mathbf{V}$

$$\min_{\mathbf{U},\mathbf{V}} \sum_{k=1}^{K} ||\mathbf{W}_{(k)} \odot (\mathbf{Y} - \mathbf{L})||_{p_k}^{p_k}$$
$$s.t \quad \mathbf{L} = \mathbf{U}\mathbf{V}^T.$$

$$L(\mathbf{U}, \mathbf{V}, \mathbf{L}, \mathbf{Y}, \rho) = \sum_{k=1}^{K} ||\mathbf{W}_{(k)} \odot (\mathbf{Y} - \mathbf{L})||_{p_k}^{p_k} + \langle \mathbf{\Lambda}, \mathbf{L} - \mathbf{U}\mathbf{V}^T \rangle + \frac{\rho}{2} ||\mathbf{L} - \mathbf{U}\mathbf{V}^T||_F^2$$

Selection of λ

$$\max_{\mathbf{\Theta}} \left\{ l_P^C(\mathbf{\Theta}) = l^C(\mathbf{\Theta}) - P(\boldsymbol{\pi}; \lambda) \right\}$$

$$P(\boldsymbol{\pi}; \lambda) = n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}$$

$$\mathrm{BIC}(\lambda) = \sum_{i,j \in \Omega} \log \{ \sum_{k=1}^{\hat{K}} \hat{\pi}_k f_k(e_{ij}; \hat{\eta}_k) \} - \frac{1}{2} (\sum_{k=1}^{\hat{K}} D_k) \log |\Omega|$$

Experiments

- Synthetic data

$$\mathbf{Y}_{gt} = \mathbf{U}_{gt}\mathbf{V}_{gt}^T$$

where $\mathbf{U}_{gt} \in \mathcal{R}^{40\times 4}$ and $\mathbf{V}_{gt} \in \mathcal{R}^{20\times 4}$

- Settings - 20% missing entries
  - 80% Gaussian noise - $\mathcal{N}(0, 0.04)$
  - 10% Sparse noise – Uniformly distributed in $[-20,20]$
  - 80% EP noise – $EP(0, 0.2^p p), p = 0.2$
  - Mixture noise
    - 20% Gaussian noise – $\mathcal{N}(0, 0.04)$
    - 40% Gaussian noise – $\mathcal{N}(0, 0.01)$
    - 20% Sparse noise - Uniformly distributed in $[-5,5]$

Synthetic Data

$$C1 = ||\mathbf{W} \odot (\mathbf{Y}_{no} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T)||_1$$

$$C2 = ||\mathbf{W} \odot (\mathbf{Y}_{no} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T)||_2$$

$$C3 = ||\mathbf{Y}_{gt} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T||_1$$

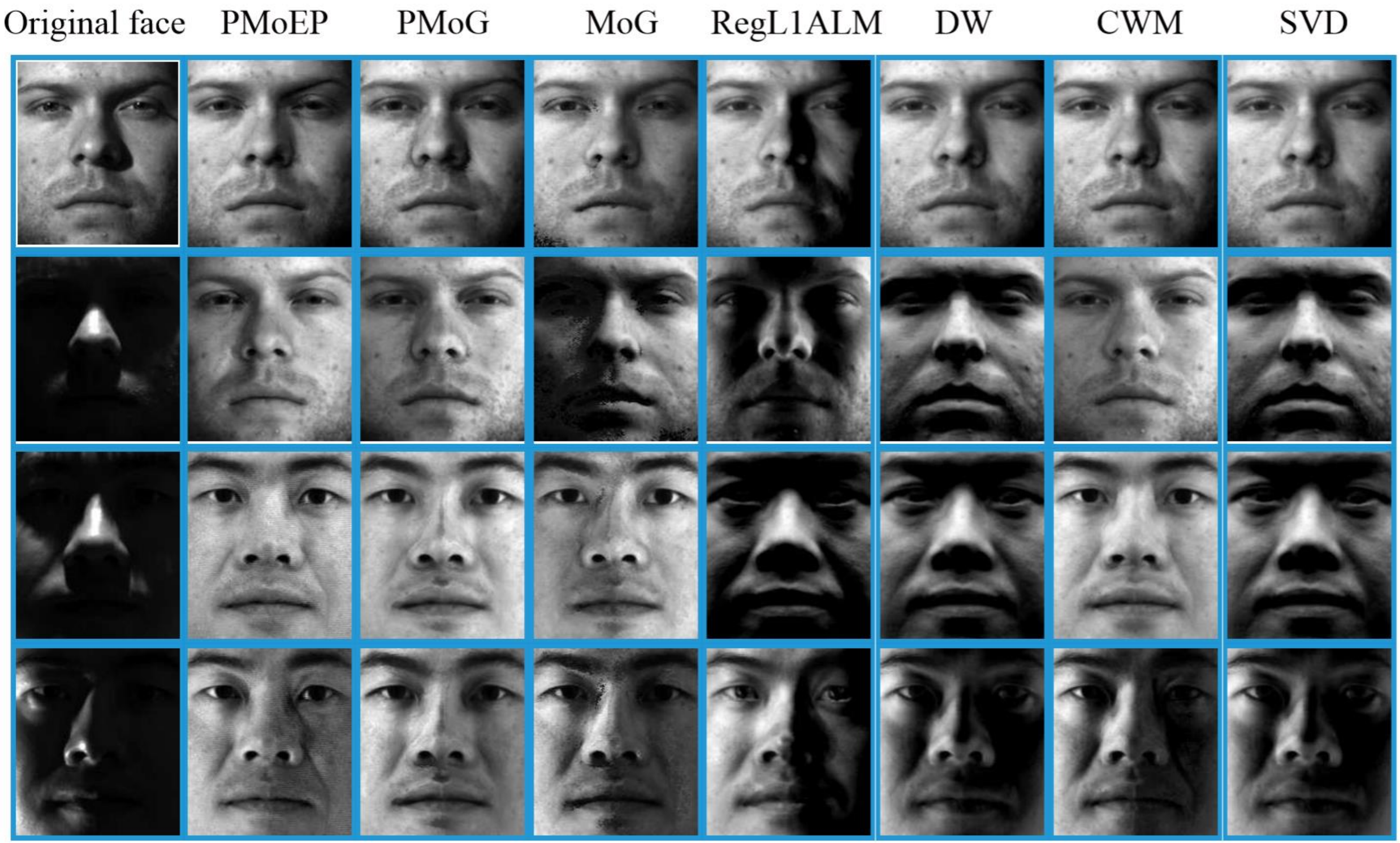
$$C4 = ||\mathbf{Y}_{gt} - \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T||_2$$

$$C5 = subspace(\mathbf{U}_{gt}, \tilde{\mathbf{U}})$$

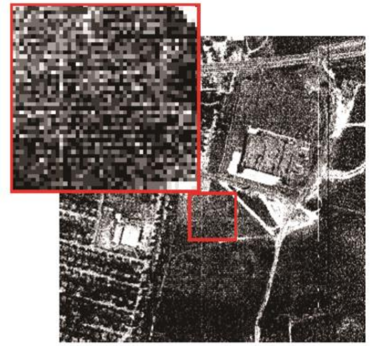
$$C6 = subspace(\mathbf{V}_{gt}, \tilde{\mathbf{V}})$$

	PMoEP	PMoG	MoG [19]	CWM [20]	DW [24]	RegL1ALM [37]
			Sparse Noise			
C1	8.32e+2	7.98e+2	8.03e+2	**7.97e+2**	1.05e+3	8.63e+2
C2	1.07e+4	1.06e+4	1.08e+4	9.97e+3	**4.79e+3**	5.96e+3
C3	3.23e+1	**2.13e-12**	1.57e+1	4.10e+2	5.62e+14	1.19e+6
C4	6.73e+1	**2.42e-5**	3.45e+1	9.69e+1	1.30e+7	4.26e+3
C5	1.40e-1	**3.85e-8**	3.30e-2	3.93e-1	1.47e+1	1.49e+1
C6	6.16e-2	**2.33e-8**	1.96e-6	1.13e-1	1.44e+1	1.45e+1
			Gaussian Noise			
C1	8.08e+1	7.83e+1	8.20e+1	7.71e+1	8.16e+1	**7.35e+1**
C2	**1.65e+1**	2.41e+1	1.67e+1	2.17e+1	1.67e+1	2.14e+1
C3	**1.31e+1**	2.48e+1	1.33e+1	2.24e+1	1.32e+1	2.03e+1
C4	7.89e+1	1.06e+2	7.90e+1	9.95e+1	**7.86e+1**	9.74e+1
C5	**8.54e-2**	1.17e-1	9.24e-2	1.22e-1	8.67e-2	1.10e-1
C6	5.82e-2	8.43e-2	6.30e-2	9.96e-2	**5.77e-2**	7.06e-2
			EP Noise			
C1	3.64e+2	3.19e+2	**3.18e+2**	3.30e+2	4.33e+2	3.53e+2
C2	1.29e+3	1.25e+3	1.02e+3	1.35e+3	**6.49e+2**	8.75e+2
C3	**1.58e+2**	3.28e+3	6.98e+4	1.99e+2	1.04e+7	6.66e+4
C4	**2.40e+2**	2.65e+2	3.52e+2	2.47e+2	2.57e+3	8.22e+2
C5	**3.02e-1**	4.21e-1	4.19e-1	3.62e-1	1.08e+1	1.14e+1
C6	**2.02e-1**	2.87e-1	3.28e-1	2.53e-1	9.05e-1	1.01e+1
			Mixture Noise			
C1	4.40e+2	4.38e+2	4.27e+2	4.31e+2	5.18e+2	**4.26e+2**
C2	1.27e+3	1.33e+3	1.28e+3	1.10e+3	**8.26e+2**	1.12e+3
C3	**1.42e+2**	1.90e+2	1.88e+2	3.77e+2	2.16e+8	1.45e+4
C4	**1.68e+2**	1.91e+2	1.84e+2	3.15e+2	7.08e+3	5.08e+2
C5	**3.47e-1**	3.90e-1	3.93e-1	5.41e-1	8.44e-1	7.96e-1
C6	1.79e-1	1.78e-1	**1.62e-1**	3.91e-1	7.46e-1	6.17e-1
Time(s)	2.34	0.28	0.46	**0.062**	0.352	0.11

Face modelling



Face modelling



(a) Original HSI

(b) SVD

(c) RegL1ALM

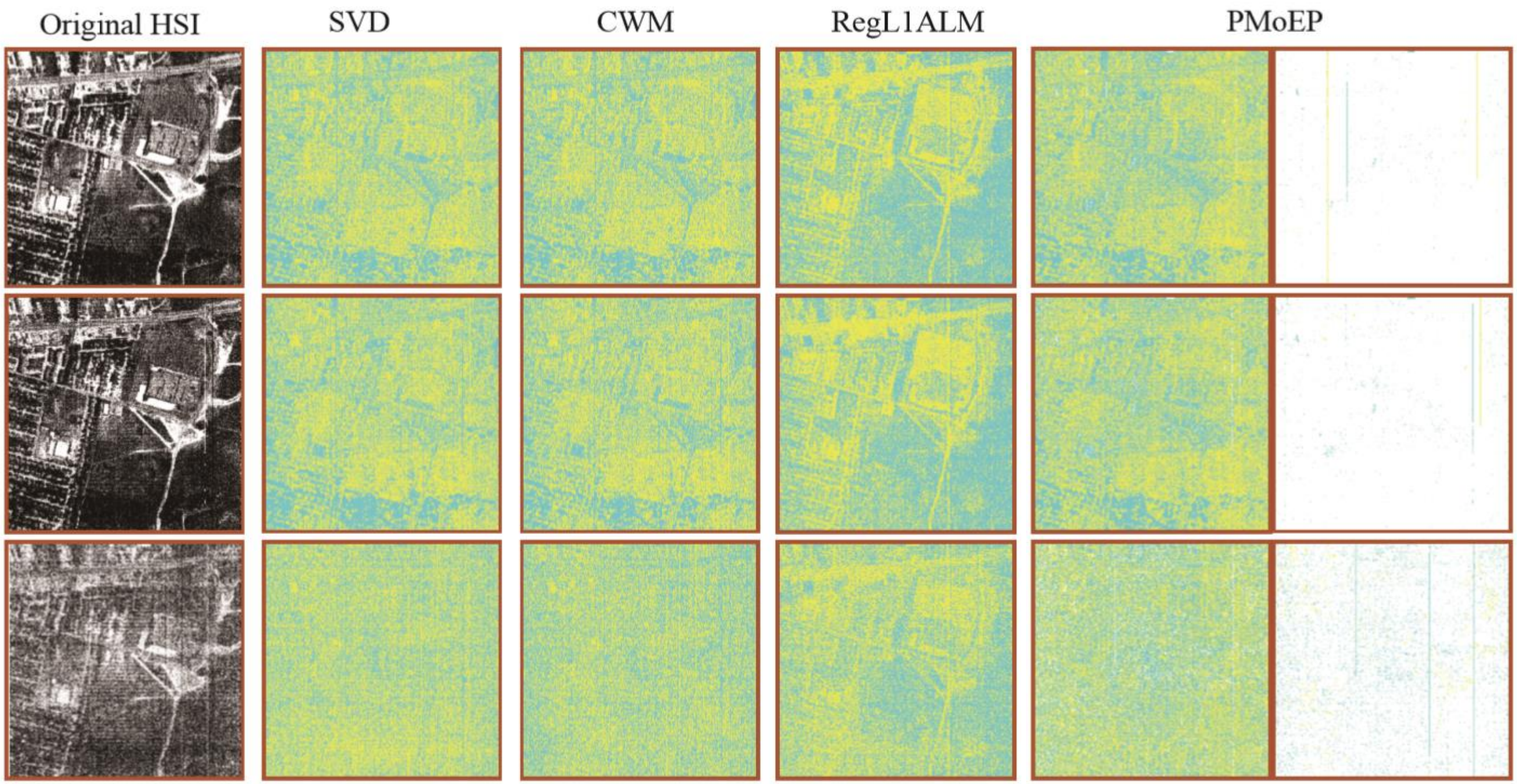
(d) CWM

(e) MoG

(f) PMoG

(g) PMoEP

Face modelling



Face modelling