Robust Matrix Factorization with Unknown Noise

作者：
Deyu Meng Xi’ an Jiaotong University
Fernando De la Torre Carnegie Mellon University

一.工作动机

许多计算机视觉问题可以被当成是从高维视觉数据找到低维子空间，一种估计低秩子空间的方法是因式分解，最小化观测数据矩阵和双线性因式分解的损失函数。最常见的损失函数由L2损失和L1损失，分别对应高斯分布和拉普拉斯分布的噪音。真实数据的噪音不单纯是高斯或拉普拉斯， 本文提出Mixture of Gaussians (MoG) noise model，适用于所有连续型的分布，可以更好的模拟噪音。

二.模型

(Low Rank Matrix Factorization)LRMF问题数学形式，最小化LRMF error

$$\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{W}\odot(\mathbf{X}-\mathbf{U}\mathbf{V}^T)\right\|_{L_p}, \qquad (1)$$

where $\mathbf{U}\in\Re^{d\times r}$ and $\mathbf{V}\in\Re^{n\times r}$ are low-dimensional matrices ($r<\min(d,n)$), $\mathbf{W}$ is the indicator matrix of the same size as $\mathbf{X}$. $w_{ij}=0$ if x_{ij} is missing and 1 otherwise. $\|\cdot\|_{L_p}$ denotes an L_p norm, and most popular approaches use L_2 and L_1 norm.

1. The subspaceMoG model

在矩阵X里面，每个小x可以写为

ε表示噪音

$$x_{ij}=(\mathbf{u}^i)^T\mathbf{v}^j+\varepsilon_{ij},$$

L1和L2范式指的是认为噪音服从拉普拉斯和最大似然分布，对噪音进行最大似然估计。MoG模型可以看做任何连续型分布的近似。假设ε服从MoG，πk代表混合比例

$$p(\varepsilon)\sim\sum_{k=1}^{K}\pi_k\mathcal{N}(\varepsilon|0,\sigma_k^2),$$

由小x的式子可以得到x的概率为，接着可以写出数据集的似然函数

$$p(x_{ij}|\mathbf{u}^i,\mathbf{v}^j,\mathbf{\Pi},\mathbf{\Sigma}) = \sum_{k=1}^{K}\pi_k p(x_{ij}|k),$$

where $p(x_{ij}|k)=\mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\sigma_k^2)$, $\mathbf{\Pi}=\{\pi_1,\pi_2,\cdots,\pi_K\}$, and $\mathbf{\Sigma}=\{\sigma_1,\sigma_2,\cdots,\sigma_K\}$. The likelihood of $\mathbf{X}$ can then be written as

$$p(\mathbf{X}|\mathbf{U},\mathbf{V},\mathbf{\Pi},\mathbf{\Sigma}) = \prod_{i,j\in\Omega} p(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\mathbf{\Pi},\mathbf{\Sigma})$$
$$= \prod_{i,j\in\Omega}\sum_{k=1}^{K}\pi_k\mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\sigma_k^2),$$

where Ω is the index set of the non-missing entries in $\mathbf{X}$.

有了似然函数，从4个参数最大化似然函数，就可以得到一个符合数据集的分解矩阵，取对数，问题转化为

$$\max_{\mathbf{U},\mathbf{V},\mathbf{\Pi},\mathbf{\Sigma}} \mathcal{L}(\mathbf{U},\mathbf{V},\mathbf{\Pi},\mathbf{\Sigma})$$
$$= \sum_{i,j\in\Omega}\log\sum_{k=1}^{K}\pi_k\mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\sigma_k^2).$$

2. EM algorithm

EM算法，估计参数最大化似然函数

E Step：假设隐变量z，表示噪音属于哪一个组分，z的条件均值 $z_{ijk} \in \{0,1\}$ and $\sum_{k=1}^{K} z_{ijk} = 1,$

$$E(z_{ijk}) = \gamma_{ijk} = \frac{\pi_k\mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\sigma_k^2)}{\sum_{k=1}^{K}\pi_k\mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j,\sigma_k^2)}.$$

最完整的似然函数，是整个噪音的分布添加幂次γ，然后取对数，代入高斯分布式子可以得到下式，问题转化为最大化下式，可以对分开的项单独进行最大化

$$E_{\mathbf{Z}}\, p(\mathbf{X},\mathbf{Z}\|\mathbf{U},\mathbf{V},\mathbf{\Pi},\mathbf{\Sigma}) =$$
$$\sum_{i,j\in\Omega}\sum_{k=1}^{K}\gamma_{ijk}\left(\log\pi_k - \log\sqrt{2\pi}\sigma_k - \frac{(x_{ij}-(\mathbf{u}^i)^T\mathbf{v}^j)^2}{2\sigma_k^2}\right).$$

M Step：maximation

Update $\mathbf{\Pi},\mathbf{\Sigma}$: Closed-form updates for the MoG parameters (for $k = 1, 2, \cdots, K$) are [12]:

$$N_k = \sum_{i,j}\gamma_{ijk},\quad \pi_k = \frac{N_k}{N},$$
$$\sigma_k^2 = \frac{1}{N_k}\sum_{i,j}\gamma_{ijk}(x_{ij}-(\mathbf{u}^i)^T\mathbf{v}^j)^2. \qquad (6)$$

Update $\mathbf{U},\mathbf{V}$: The components of Eq. (5) related to $\mathbf{U}$ and $\mathbf{V}$ can then be re-written as follows:

$$\sum_{i,j\in\Omega}\sum_{k=1}^{K}\gamma_{ijk}\left(-\frac{(x_{ij}-(\mathbf{u}^i)^T\mathbf{v}^j)^2}{2\sigma_k^2}\right)$$
$$= -\sum_{i,j\in\Omega}\left(\sum_{k=1}^{K}\frac{\gamma_{ijk}}{2\sigma_k^2}\right)(x_{ij}-(\mathbf{u}^i)^T\mathbf{v}^j)^2$$
$$= -\left\|\mathbf{W}\odot(\mathbf{X}-\mathbf{U}\mathbf{V}^T)\right\|_{L_2}, \qquad (7)$$

Alternated Least Squares

Algorithm 1: MoG algorithm for LRMF

Input: $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \in R^{d \times n}$, index set Ω of non-missing entries of $\mathbf{X}$

Output: $\mathbf{U}, \mathbf{V}$

Randomly initialize $\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{U}, \mathbf{V}$, MoG number K, small threshold ε.

repeat

(*E Step*): Evaluate γ_{ijk} for $i = 1, 2, \cdots, d$; $j = 1, 2, \cdots, n$; $k = 1, 2, \cdots, K$ by Eq. (4).

(*M Step for* $\mathbf{\Pi}, \mathbf{\Sigma}$): Evaluate π_k, σ_k^2 for $k = 1, 2, \cdots, K$ by Eq. (6).

(*M Step for* $\mathbf{U}, \mathbf{V}$): Evaluate $\mathbf{U}, \mathbf{V}$ by solving $\min_{\mathbf{U},\mathbf{V}} \left\| \mathbf{W} \odot (\mathbf{X} - \mathbf{U}\mathbf{V}^T) \right\|_{L_2}$ through ALS, where $\mathbf{W}$ is calculated by Eq. (8).

(*Automatic K tuning*): If $\frac{|\sigma_i^2 - \sigma_j^2|}{\sigma_i^2 + \sigma_j^2} < \varepsilon$ for some i, j, then combine the i^{th} and j^{th} Gaussian components into a unique Gaussian by letting $\pi_i = \pi_i + \pi_j$, $\sigma_i^2 = (n_i\sigma_i^2 + n_j\sigma_j^2)/(n_i + n_j)$, where n_i is the element number in i^{th} Gaussian component, and removing the j^{th} Gaussian parameters from $\mathbf{\Pi}, \mathbf{\Sigma}$. Let $K = K - 1$.

until *converge*;

三.实验

1. Synthetic experiments

$$E1 = \left\| \mathbf{W} \odot (\mathbf{X}_{no} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_1}, E2 = \left\| \mathbf{W} \odot (\mathbf{X}_{no} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_2},$$

$$E3 = \left\| \mathbf{X}_{gt} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T \right\|_{L_1}, E4 = \left\| \mathbf{X}_{gt} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T \right\|_{L_2},$$

$$E5 = subspace(\mathbf{U}_{gt}, \widetilde{\mathbf{U}}), E6 = subspace(\mathbf{V}_{gt}, \widetilde{\mathbf{V}}),$$

where $\mathbf{U}, \mathbf{V}$ are the outputs of the corresponding LRMF method, and $subspace(\mathbf{U_1}, \mathbf{U_2})$ denotes the angle between subspaces spanned by the columns of $\mathbf{U_1}$ and $\mathbf{U_2}$.

	MoG	IRLS[11]	WLRA[35]	DN[6]	ALP[20]	L1Wiberg[14]	RegL1ALM[45]	CWM[27]	NN-Robust PCA[7]
					No Noise (log values)				
$E1$	*−18.5*	**−26.5**	−2.17	−11.4	0.587	−8.73	−12.6	−0.222	−2.86
$E2$	*−42.2*	**−56.8**	−9.38	−27.6	−4.27	−21.1	−30.0	−4.83	−10.8
$E3$	*−17.9*	**−25.9**	−1.50	−11.0	1.03	−8.44	−11.9	0.324	−2.20
$E4$	*−40.5*	**−55.2**	−7.06	−27.3	−3.37	−20.7	−27.9	−3.65	−8.73
$E5$	*−23.8*	**−32.5**	−7.05	−18.9	−5.28	−14.6	−17.4	−5.35	−7.75
$E6$	*−23.9*	**−31.9**	−8.15	−19.4	−6.18	−16.1	−18.4	−5.95	−8.83
					Gaussian Noise				
$E1$	40.0	40.0	40.0	40.0	37.5	**35.7**	**35.7**	39.1	40.0
$E2$	**3.99**	**3.99**	**3.99**	**3.99**	5.20	5.08	5.09	5.59	**3.99**
$E3$	**39.3**	**39.3**	**39.3**	**39.3**	49.3	47.95	48.02	51.9	**39.3**
$E4$	**3.29**	**3.29**	**3.29**	**3.29**	5.31	4.91	4.93	6.38	**3.29**
$E5$	**0.0455**	**0.0455**	0.0456	**0.0455**	0.0561	0.0541	0.0545	0.0636	**0.0455**
$E6$	**0.0295**	**0.0295**	**0.0295**	**0.0295**	0.0395	0.0368	0.0367	0.0466	**0.0295**
					Sparse Noise				
$E1$	*400.6*	519.1	518.5	519.1	403.6	**395.1**	425.8	425.7	523.1
$E2$	1317.7	*827.5*	*827.5*	**827.3**	1174.8	1270.8	1176.1	1125.1	834.7
$E3$	**54.1**	624.1	623.2	624.4	159.9	*67.4*	446.4	278.0	628.0
$E4$	*100.9*	991.1	1002.3	995.4	171.2	**99.4**	8984.9	342.4	976.05
$E5$	*0.289*	0.741	0.740	0.742	0.347	**0.279**	0.978	0.471	0.733
$E6$	**0.0169**	0.583	0.581	0.584	0.164	*0.0284*	0.698	0.338	0.587
					Mixture Noise				
$E1$	419.7	516.9	516.9	516.8	*412.9*	**404.5**	417.9	430.8	520.5
$E2$	1274.3	**829.0**	**829.0**	**829.0**	1119.1	1147.1	1124.5	1120.4	836.3
$E3$	**149.5**	616.1	615.7	616.1	242.1	*192.4*	375.4	291.6	618.4
$E4$	**189.1**	956.1	951.7	956.2	276.5	*213.5*	4245.4	336.6	955.9
$E5$	**0.374**	0.692	0.691	0.692	0.427	*0.377*	0.696	0.461	0.701
$E6$	**0.155**	0.579	0.579	0.579	0.230	*0.175*	0.530	0.346	0.589

Table 1. Six measures of error for the synthetic example with different noise models. The best and the second best results in each experiment are highlighted in bold and italic, respectively.

performs best or the second best in all experiments in estimating a better subspace from noisy data (measurements E3-E6).

2. SFM (Structure From Motion) experiments

$$E1 = \left\| \mathbf{W} \odot (\mathbf{X}_{no} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_1}, E2 = \left\| \mathbf{W} \odot (\mathbf{X}_{no} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_2},$$

$$E3 = \left\| \mathbf{W} \odot (\mathbf{X}_{gr} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_1}, E4 = \left\| \mathbf{W} \odot (\mathbf{X}_{gr} - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T) \right\|_{L_2},$$

	MoG	IRLS[11]	WLRA[35]	DN[6]	ALP[20]	RegL1ALM[45]	CWM[27]	NN-Robust PCA[7]
					No Noise			
$E1$	*0.442*	1.83	8.24	0.490	4.85	**0.291**	7.71	4.98
$E2$	*1.13*	3.18	12.1	**1.12**	12.9	1.47	26.4	7.96
$E3$	*0.442*	1.83	8.24	0.490	4.85	**0.291**	7.71	4.98
$E4$	*1.13*	3.18	12.1	**1.12**	12.9	1.47	26.4	7.96
					Gaussian Noise			
$E1$	*6.70*	7.03	11.2	6.73	8.48	**6.14**	12.2	9.05
$E2$	*8.48*	8.93	14.9	**8.43**	15.7	9.31	21.9	12.2
$E3$	*4.55*	5.01	9.41	**4.49**	8.05	5.39	11.6	7.04
$E4$	*5.87*	6.43	13.2	**5.85**	14.2	6.91	20.7	10.2
					Sparse Noise			
$E1$	*3.24*	4.95	10.48	4.53	6.42	**2.85**	10.3	7.44
$E2$	9.31	*8.27*	14.90	**7.90**	17.7	9.22	19.6	11.5
$E3$	*1.29*	4.08	9.34	3.58	4.91	**0.524**	9.02	6.41
$E4$	*4.28*	6.13	13.09	5.49	15.8	**1.98**	18.0	9.68
					Mixture Noise			
$E1$	*8.27*	8.70	12.8	8.42	10.6	**7.67**	13.1	10.0
$E2$	*11.2*	11.4	16.9	**10.99**	18.2	12.3	23.8	13.3
$E3$	**5.82**	5.97	9.99	5.83	9.19	6.31	11.7	7.19
$E4$	**7.81**	7.88	13.8	7.82	15.5	8.44	21.5	9.83

Table 2. Performance comparison of the competing methods in the SFM experiments. The best and the second best results in each experiment are highlighted in bold and italic, respectively.

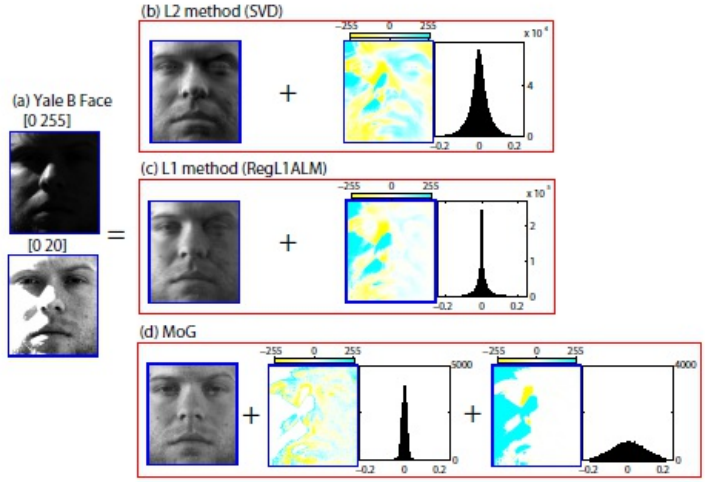
always has the best or the second best performance in the error E3-E6

3.Face modeling experiments



Figure 2. From left to right: original face images, faces reconstructed by MoG, Robust PCA (IRLS), SVD, RegL1ALM, PCAL1, CWM and nuclear-norm based Robust PCA, respectively.

performs better over the faces with large dark regions



4.Background subtraction experiments

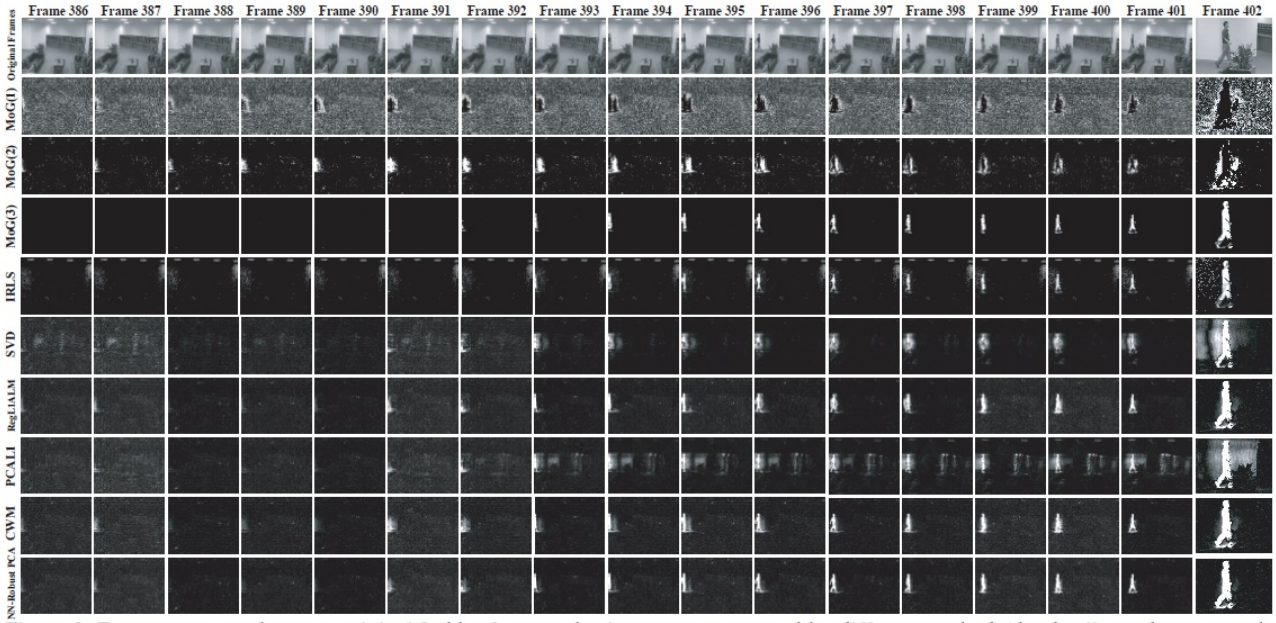


Figure 3. From top row to bottom: original Lobby frames, absolute errors computed by different methods(the details are better seen by zooming on a computer screen). The moving object region in Frame 402 is enlarged for better visualization.

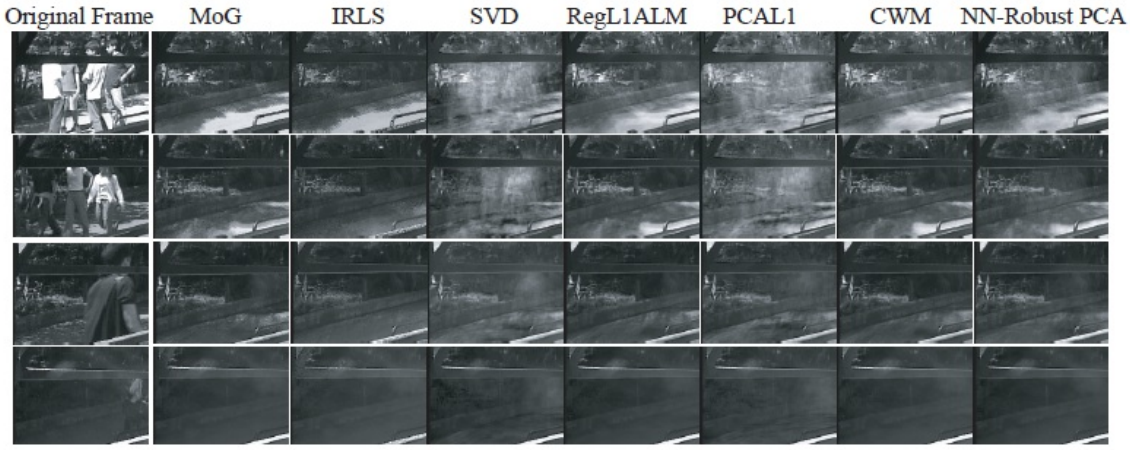


Figure 4. From left to right: original frames, reconstructed backgrounds computed by different methods.

四.总结

假设噪音服从MoG模型，自动从子空间高斯模型里面学习参数u和v，可以同时适应高斯和拉普拉斯噪音，同时对于混合噪音拟合效果最好