Robust Online Matrix Completion with Gaussian Mixture Model

Chunsheng Liu, Chunlei Chen, Hong Shan, Bin Wang

ICASSP 2020

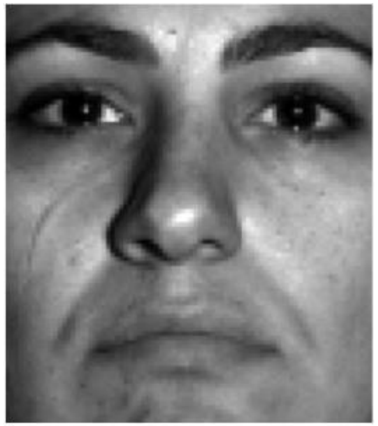
Session Title: [Dictionary Learning, Representation Learning and Matrix Completion] MLSP Poster Session 1

Outline

- Background
- Problem Formulation and Modeling
- Solution
- Main Results
- Conclusion

Background

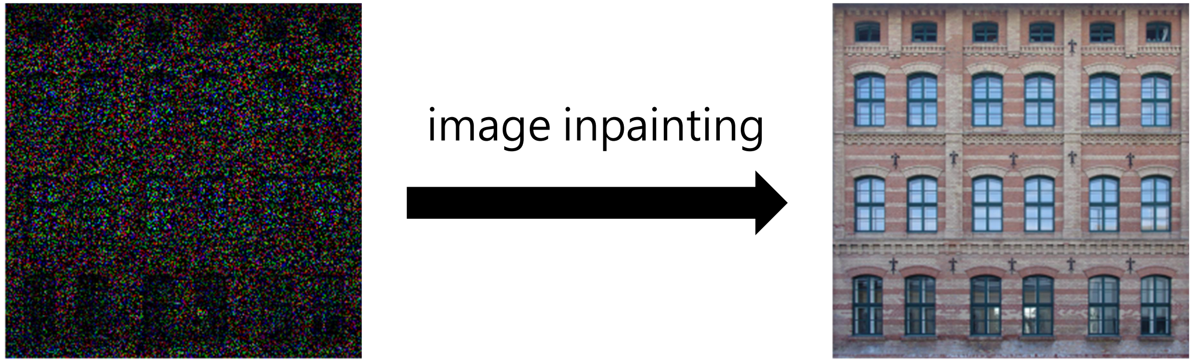
Low-rank structure is common in practice.

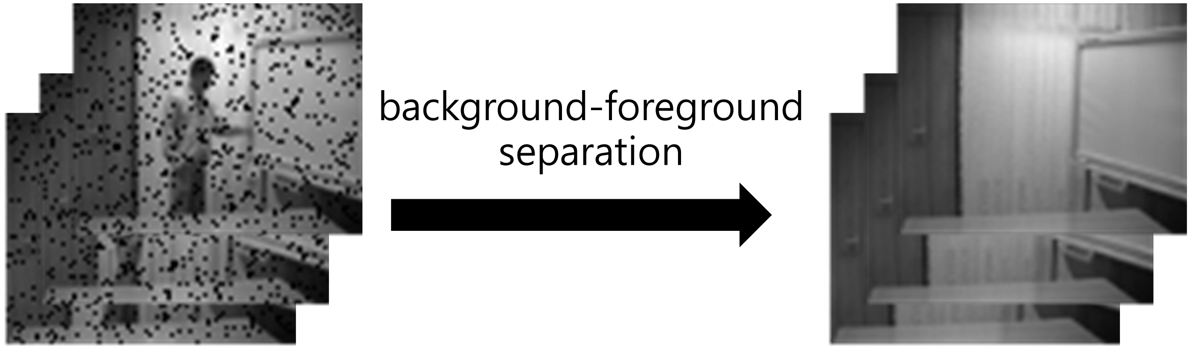


Learning by using the underlying low-rank structure of data is important.

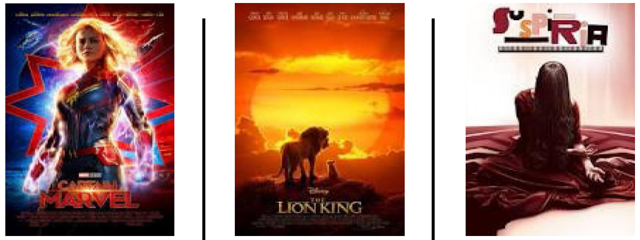
Background

Utilizing the low-rank property of data, **matrix completion** attempts to recover the data when only some of elements are observed.





Movies



Users

User 1	3	?	?
User 2	?	2	4
User 3	?	3	?
User 4	4	?	4

Recommender system (The Netflix Prize problem)

Exiting work

In modern context, data are collected in a streaming fashion.

real-time process of high dimensional data ⟷ **mismatch** ⟷ limited computational resources

⬇

matrix completion with batch setting ⟷ **extend** ⟷ matrix completion with streaming setting (online matrix completion)

Exiting work

Mathematically, at each time t, the online matrix completion (MC) problem can be formulated as:

$$\arg\min_{\ell_t, s_t} f(\ell_t, s_t) + \lambda r(s_t).$$

smooth loss function (f)

the regularization function (r)

$\ell_t \in \mathbb{R}^n$ —— underlying data vector arriving at time t

$s_t \in \mathbb{R}^n$ —— noise data vector arriving at time t

Exiting work

ICASSP2020 Barcelona

Broadly speaking, there are roughly two perspectives from which researchers have developed and studied online MC methods, categorized by the choice of r.

Table 1 : The two perspectives of exiting online MC methods (categorized by the choice of r).

	Method	Model
L_1-norm	Modified-PCP	$\min_{\ell_t,s_t} \lambda\|\|s_t\|\|_1 + \|\|\left(I - U_tU_t^\top\right)\|\|_*$
	ReProCS	$\min_{\ell_t,s_t} \|\|s_t\|\|_1 \ \ s.t.\|\|m_t - \left(I - U_tU_t^\top\right)s_t\|\| < \eta$
	NORST	$\min_{\ell_t,s_t} \|\|s_t\|\|_1 \ \ s.t.\|\|m_t - \left(I - U_tU_t^\top\right)s_t\|\| < \eta$
L_2-norm	PETRELS	$\min_{\ell_t,s_t} \|\|P_\Omega\left(m_t - U_{t-1}s_t\right)\|\|_2{}^2$
	GROUSE	$\min_{\ell_t,s_t} \|\|P_{\Omega_t}\left(s_t - U_tU_t^\top s_t\right)\|\|_2{}^2$

U_t is low-rank subspace, $m_t = \ell + s_t$

Motivation

- The fixed loss term is adopted in these models (implicitly assumes that noises involved in the data follow a fixed probability distribution).

- Such assumption deviates from the real scenarios where noise is always time-varying [Allili et al. 2007].

- **Our scenario:** study the online MC problem with the variations in both low-rank subspace and noises.

Contribution

- Model noises as Gaussian Mixture Model (GMM) distribution rather than a fixed single noise distribution.
- Present the online MC model by employing the proposed GMM regularizer and solved it by a fast and memory-efficient algorithm.

Modeling

main idea: fit a specific GMM distribution for the noise term x_t in each newly coming observed data vector y_t .

noise data modeling: $$x_t^i \sim \sum_{k=1}^{K} \pi_t^k \mathcal{N}\left(x_t^i \middle| 0, \left(\sigma_t^k\right)^2\right).$$

Upon receiving the data vector y_t , the posterior distribution is:

observed data modeling: $$y_t^i \sim \prod_{k=1}^{K} \mathcal{N}\left(y_t^i \middle| \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right), \left(\sigma_t^k\right)^2\right)^{z_t^{ik}} .$$

latent variable $z_t^i \sim \mathcal{M}\left(z_t^i | \Pi\right)$

Problem Formulation

The online MC model with GMM regularizer (named ONCE) can be formulated as:

$$\arg\min_{\ell_t, x_t} \lambda \underline{\|x_t\|_{\mathrm{M}}} + \frac{1}{2} \left\| \mathcal{P}_{\Omega_t}(\ell_t) + x_t - y_t \right\|_2^2 .$$

the defined GMM regularizer:

$$\|x_t\|_{\mathrm{M}} = \alpha_\sigma \|x_t\|_2^2 + \beta_{\pi,\sigma} + c.$$

Solution

The above optimization problem can be solved through three stages:

- Update the GMM parameters (π_t^k, σ_t^k) utilizing EM procedure.

$$\pi_t^k = \pi_{t-1}^k - \frac{n}{\bar{n}}\left(\pi_{t-1}^k - \bar{\pi}^k\right). \qquad \left(\sigma_t^k\right)^2 = \left(\sigma_{t-1}^k\right)^2 - \frac{n^k}{\bar{n}^k}\left(\left(\sigma_{t-1}^k\right)^2 - \left(\bar{\sigma}^k\right)^2\right).$$

- Update noise vector and underlying data vector
  - **Update x_t :** $x_t = \frac{y_t - \mathcal{P}_{\Omega_t}(\ell_{t-1})}{1+2\lambda\alpha_\sigma}.$
  - **Update ℓ_t :** $\ell_t = y_t - \mathbf{I}_{\tilde{\Omega}_t^c}\left(\Phi_{\tilde{\Omega}_t^c}^\top \Phi_{\tilde{\Omega}_t^c}\right)^{-1} \Phi_{\tilde{\Omega}_t^c}^\top \Phi y_t.$

- Update the low-dimensional subspace $\mathbf{U}_t$ with the strategy in [Narayanamurthy et al. 2019]

Main Results

Experiments: online matrix completion

Evaluation metrics:

i) error ratio (ER) $\mathrm{ER} = \frac{\|\hat{\ell}_t - \ell_t\|_2}{\|\ell_t\|_2}$

ii) mean of the absolute errors (MAE) $\mathrm{MAE} = \frac{1}{n}\sum_i \left|\hat{\ell}_t^i - \ell_t^i\right|$

iii) computation time

Main Results

Experiments: online matrix completion

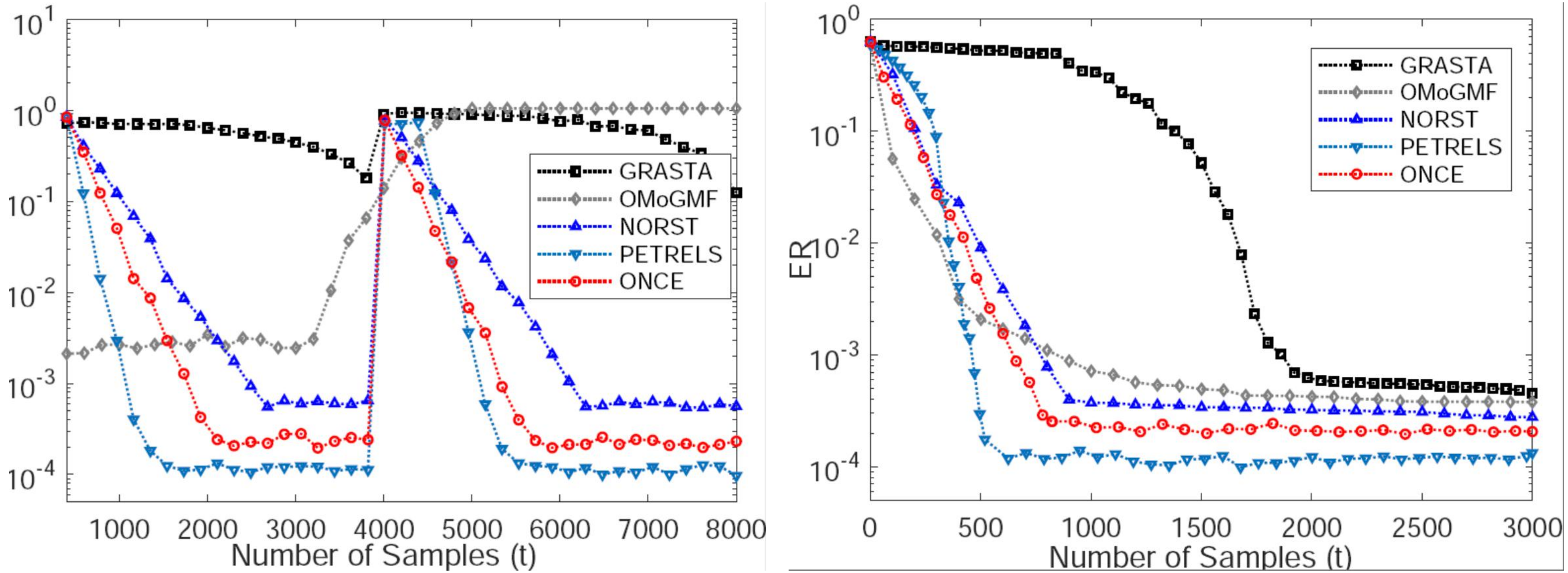


Figure 1: The comparison results of error ratios. The sampling ratio is fixed at 0.5. From left to right: ER versus number of samples on (a) synthetic data with piecewise constant subspace change; (b) synthetic data with subspace changing at each time

Main Results

Experiments: online matrix completion for traffic data

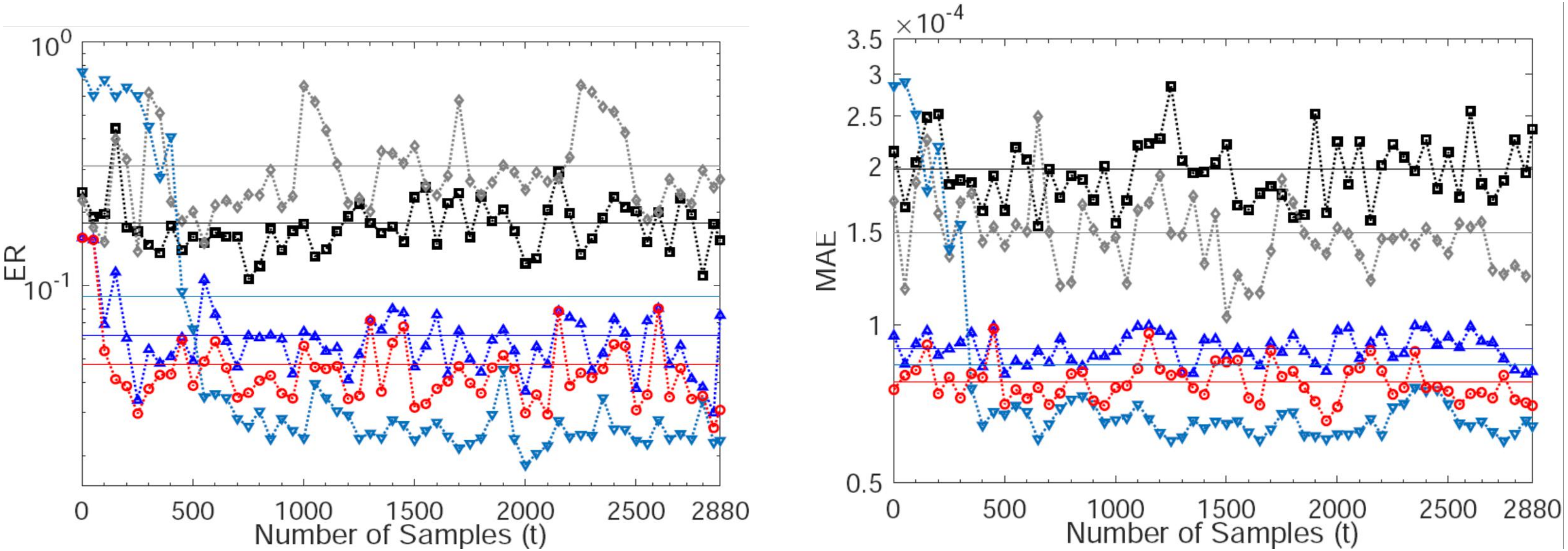
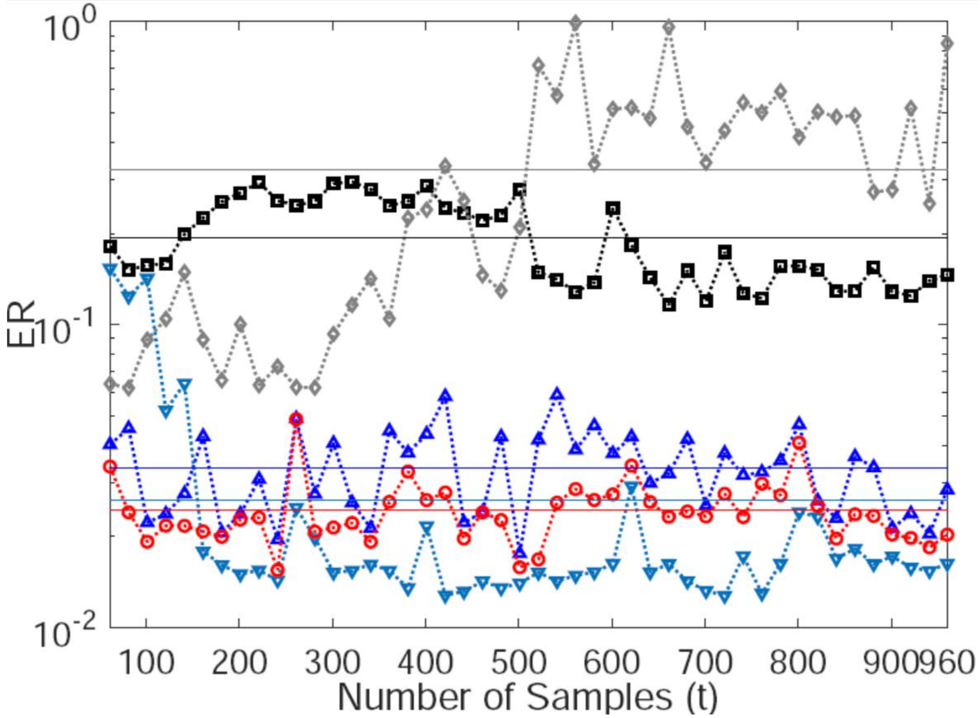


Figure 2: The comparison results on traffic data (Abilene). The sampling ratio is fixed at 0.5.

Main Results

Experiments: online matrix completion for traffic data



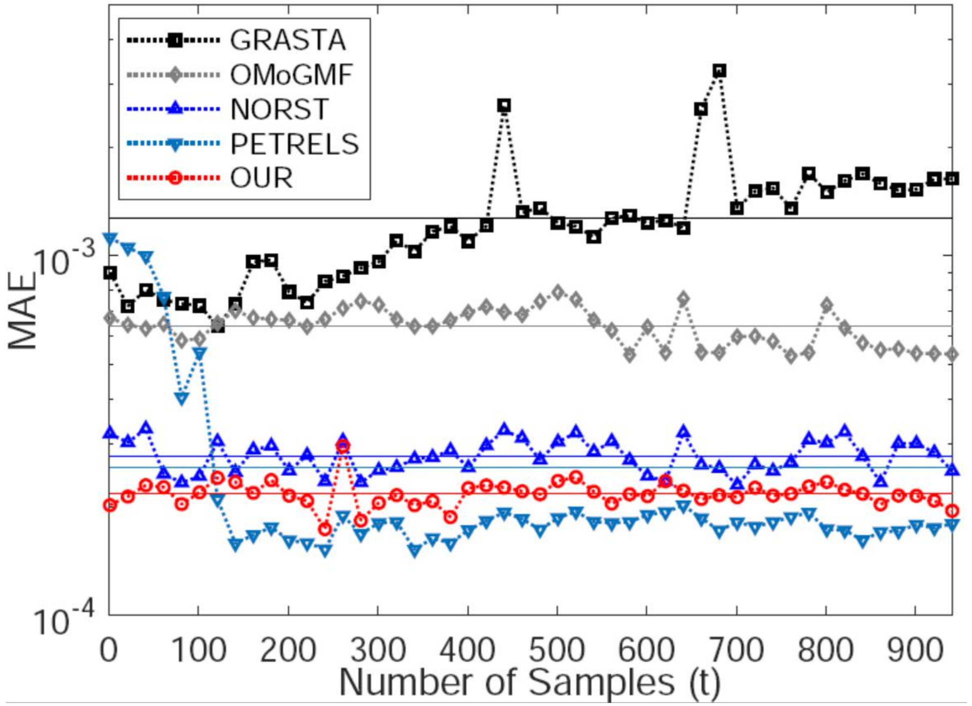


Figure 3: The comparison results on traffic data (Geant). The sampling ratio is fixed at 0.5.

Main Results

ICASSP2020 Barcelona

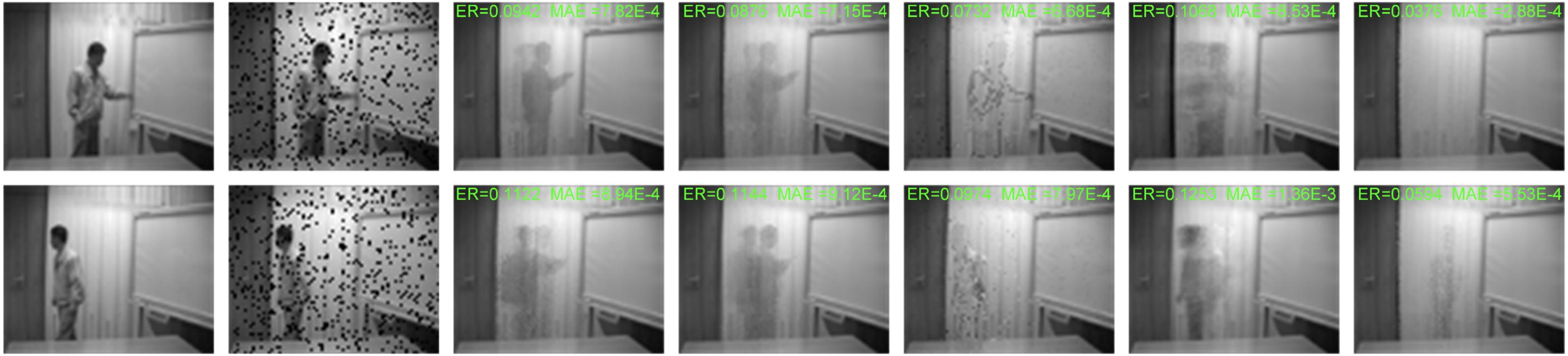
Experiments: online matrix completion

Table 2: The comparison results of MAE and computation time. The results are rounded to two decimal places, where MAEs are in 10^{-4} and times are in milliseconds.

	Synthetic data				Real traffic data			
	Piecewise		Changing		Abilene		Geant	
	MAE	time	MAE	time	MAE	time	MAE	time
GRASTA	10.90	4.38	9.02	3.93	1.98	0.94	6.37	1.25
OMoGMF	13.58	15.90	4.16	16.00	1.50	1.83	12.73	7.09
NORST	0.22	5.98	1.82	2.77	0.90	0.65	2.78	1.43
PETRELS	**0.002**	70.51	**1.47**	84.31	0.84	5.35	2.57	30.21
ONCE	0.01	6.62	1.58	3.10	**0.08**	1.13	**2.18**	2.05

Main Results

Experiments: online matrix completion for video data



Original | Corrupted | GRASTA (62.95) | OMoGMF (239.94) | NORST (79.57) | PETRELS (1689.66) | ONCE (85.70)

Figure 4: The comparison results on video data. The missing ratio is fixed at 0.1.

Conclusion

ICASSP2020 Barcelona

- we presented a new online MC method aiming to high accuracy with low time and memory overheads.
- In particular, by employing the GMM noise distribution, our model is sufficiently robust to adapt real-time noise variations.

Reference

Mohand Saıd Allili, Nizar Bouguila, and Djemel Ziou, "A robust video foreground segmentation by using generalized gaussian mixture modeling," in Fourth Canadian Conference on Computer and Robot Vision (CRV 2007), 28-30 May, Montreal, Quebec, Canada, 2007,pp. 503–509.

Yuejie Chi, Yonina C. Eldar, and A. Robert Calderbank, "PETRELS: parallel subspace estimation and tracking by recursive least squares from partial observations," IEEE Trans. Signal Processing, vol. 61, no. 23, pp. 5947–5959, 2013.

Praneeth Narayanamurthy, Vahid Daneshpajooh, and Namrata Vaswani, "Provable subspace tracking from missing data and matrix completion," IEEE Trans. Signal Processing, vol. 67, no. 16, pp. 4245–4260, 2019.