Supplementary Material for “Robust Online Matrix Factorization for Dynamic Background Subtraction”

Hongwei Yong, Deyu Meng, Wangmeng Zuo, Lei Zhang

Abstract—In this supplementary material, we provide proofs of several theoretical results, give inference details of the proposed algorithm, and introduce parameter setting strategies in our experiments presented in the maintext.

Index Terms—proofs to Theorems, parameter settings.

A1. Proof to the theorem on relationship between conjugate prior and KL divergence

We first present the proof to the theorem on the relationship between conjugate prior and KL divergence, as described in Section 3.4 of the maintext.

Theorem 1 *If a distribution $p(\mathbf{x}|\boldsymbol{\theta})$ belongs to the full exponential family, which means it has the following form:*

$$p(\mathbf{x}|\boldsymbol{\theta}) = \eta(\boldsymbol{\theta})exp(\boldsymbol{\theta}^T\boldsymbol{\phi}(\mathbf{x})),$$

and its conjugate prior follows:

$$p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma) = f(\boldsymbol{\mathcal{X}},\gamma)\eta(\boldsymbol{\theta})^\gamma exp(\gamma\boldsymbol{\theta}^T\boldsymbol{\mathcal{X}}),$$

then,

$$ln\ p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma) = -\gamma D_{KL}(p(\mathbf{x}|\boldsymbol{\theta}^*)||p(\mathbf{x}|\boldsymbol{\theta})) + C,$$

where $\boldsymbol{\theta}^* = arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma)$ and C is a constant independent of $\boldsymbol{\theta}$.

Proof. It can be deduced that:

$$\begin{aligned}
&D_{KL}(p(\mathbf{x}|\boldsymbol{\theta}^*)||p(\mathbf{x}|\boldsymbol{\theta}))\\
&= \int p(\mathbf{x}|\boldsymbol{\theta}^*)\ln\frac{p(\mathbf{x}|\boldsymbol{\theta}^*)}{p(\mathbf{x}|\boldsymbol{\theta})}d\mathbf{x}\\
&= -\int p(\mathbf{x}|\boldsymbol{\theta}^*)\ln p(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x} + C_1\\
&= -\int \eta(\boldsymbol{\theta}^*)\exp(\boldsymbol{\theta}^{*T}\boldsymbol{\phi}(\mathbf{x}))(\ln\eta(\boldsymbol{\theta}) + \boldsymbol{\theta}^T\boldsymbol{\phi}(\mathbf{x}))d\mathbf{x} + C_1\\
&= -\ln\eta(\boldsymbol{\theta})\int\eta(\boldsymbol{\theta}^*)\exp(\boldsymbol{\theta}^{*T}\boldsymbol{\phi}(\mathbf{x}))d\mathbf{x}\\
&\quad -\int\eta(\boldsymbol{\theta}^*)\exp(\boldsymbol{\theta}^{*T}\boldsymbol{\phi}(\mathbf{x}))\boldsymbol{\theta}^T\boldsymbol{\phi}(\mathbf{x})d\mathbf{x} + C_1\\
&= -\ln\eta(\boldsymbol{\theta}) + \boldsymbol{\theta}^T\frac{\nabla\eta(\boldsymbol{\theta}^*)}{\eta(\boldsymbol{\theta}^*)} + C_1,
\end{aligned} \tag{1}$$

where $C_1 = \int p(\mathbf{x}|\boldsymbol{\theta}^*)\ln p(\mathbf{x}|\boldsymbol{\theta}^*)d\mathbf{x}$.

Since $\boldsymbol{\theta}^* = arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma)$, we have

$$\nabla_{\boldsymbol{\theta}}\ln p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \gamma(\frac{\nabla\eta(\boldsymbol{\theta})}{\eta(\boldsymbol{\theta})} + \boldsymbol{\mathcal{X}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = 0.$$

Then we can obtain

$$\frac{\nabla\eta(\boldsymbol{\theta}^*)}{\eta(\boldsymbol{\theta}^*)} = -\boldsymbol{\mathcal{X}}. \tag{2}$$

Thus

$$\begin{aligned}
\ln p(\boldsymbol{\theta}|\boldsymbol{\mathcal{X}},\gamma) &= \gamma(\ln\eta(\boldsymbol{\theta}) + \boldsymbol{\theta}^T\boldsymbol{\mathcal{X}}) + C_2\\
&= \gamma(\ln\eta(\boldsymbol{\theta}) - \boldsymbol{\theta}^T\frac{\nabla\eta(\boldsymbol{\theta}^*)}{\eta(\boldsymbol{\theta}^*)}) + C_2\\
&= -\gamma D_{KL}(p(\mathbf{x}|\boldsymbol{\theta}^*)||p(\mathbf{x}|\boldsymbol{\theta})) + \gamma C_1 + C_2\\
&= -\gamma D_{KL}(p(\mathbf{x}|\boldsymbol{\theta}^*)||p(\mathbf{x}|\boldsymbol{\theta})) + C,
\end{aligned} \tag{3}$$

where $C_2 = \ln f(\boldsymbol{\mathcal{X}},\gamma)$ and $C = \gamma C_1 + C_2$, which is independent of $\boldsymbol{\theta}$.

The proof is then completed. ■

A2. Proof to relationship of the online models to the offline ones

We then provide the proof to the Theorem 2 in the main text as follows.

Theorem 2 *if N^{t-1} and ρ are set to be $(t-1)d$ and 1, respectively, then the minimization problem of (12) for $\{\mathbf{\Pi},\mathbf{\Sigma}\}$ and that of (17) for $\mathbf{U}$ are equivalent to calculating:*

$$\begin{aligned}
\{\mathbf{\Pi}^t,\mathbf{\Sigma}^t\} &= arg\max_{\mathbf{\Pi},\mathbf{\Sigma}}\sum_{j=1}^{t}\ln p(\mathbf{x}^j,\mathbf{z}^j|\mathbf{\Pi},\mathbf{\Sigma},\mathbf{v}^j,\mathbf{U}^j),\\
\mathbf{U}^t &= arg\max_{\mathbf{U}}\sum_{j=1}^{t}\ln p(\mathbf{x}^j,\mathbf{z}^j|\mathbf{\Pi}^j,\mathbf{\Sigma}^j,\mathbf{v}^j,\mathbf{U}),
\end{aligned} \tag{4}$$

respectively[1]*. Moreover, under these settings, it holds that:*

$$\begin{aligned}
&||\mathbf{\Sigma}^t - \mathbf{\Sigma}^{t-1}||_F \le O(\frac{1}{t});\quad ||\mathbf{\Pi}^t - \mathbf{\Pi}^{t-1}||_F \le O(\frac{1}{t});\\
&||\mathbf{U}^t - \mathbf{U}^{t-1}||_F \le O(\frac{1}{t}).
\end{aligned} \tag{5}$$

1. $\mathbf{z}^j$ can be considered as the $E(\mathbf{z}^j)$ solved in time j.

Proof. We firstly prove the Eq. (4) for $\{\mathbf{\Pi}, \mathbf{\Sigma}\}$ as follows:

$$\begin{aligned}
&\sum_{j=1}^{t-1} \ln p(\mathbf{x}^j, \mathbf{z}^j|\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{v}^j, \mathbf{U}^j),\\
&= \sum_{j=1}^{t-1}\sum_{i=1}^{d}\sum_{k=1}^{K} z_{ik}^j \{\ln \pi_k + \ln \mathcal{N}(x_i^t|(\mathbf{u}_i)^T\mathbf{v}^j, {\sigma_k}^2)\}\\
&= \sum_{j=1}^{t-1}\sum_{i=1}^{d}\sum_{k=1}^{K} z_{ik}^j \{\ln \pi_k - \ln \sigma_k - \frac{(x_i^j - \mathbf{u}_i^T\mathbf{v}^j)^2}{2{\sigma_k}^2}\} + C\\
&= \sum_{k=1}^{K}\{\sum_{j=1}^{t-1}\sum_{i=1}^{d} z_{ik}^j \ln \pi_k - \sum_{j=1}^{t-1}\sum_{i=1}^{d} z_{ik}^j \ln \sigma_k -\\
&\quad \frac{\sum_{j=1}^{t-1}\sum_{i=1}^{d}(x_i^j - \mathbf{u}_i^T\mathbf{v}^j)^2}{2\sigma_k^2}\} + C\\
&= \sum_{k=1}^{K}\{N_k^{t-1}\ln \pi_k - N_k^{t-1}\ln \sigma_k -\\
&\quad \frac{N_k^{t-1}\frac{1}{N_k^{t-1}}\sum_{j=1}^{t-1}\sum_{i=1}^{d} z_{ik}^j(x_i^j - \mathbf{u}_i^T\mathbf{v}^j)^2}{2\sigma_k^2}\} + C\\
&= \sum_{k=1}^{K}\{N_k^{t-1}\ln \pi_k - N_k^{t-1}\ln \sigma_k - \frac{N_k^{t-1}{\sigma_k^{t-1}}^2}{2\sigma_k^2}\} + C\\
&= \sum_{k=1}^{K} N_k^{t-1}\{\ln \pi_k - \ln \sigma_k - \frac{{\sigma_k^{t-1}}^2}{2\sigma_k^2}\} + C\\
&= N^{t-1}\sum_{k=1}^{K}\pi_k^{t-1}\ln \pi_k - \sum_{k=1}^{K} N_k^{t-1}(\frac{1}{2}\frac{{\sigma_k^{t-1}}^2}{\sigma_k^2} + \ln \sigma_k) + C\\
&= -\mathcal{R}_F^t(\mathbf{\Pi}, \mathbf{\Sigma}) + C,
\end{aligned} \tag{6}$$

where

$$\begin{aligned}
&N_k^{t-1} = \sum_{j=1}^{t-1}\sum_{i=1}^{d} z_{ik}^j, N^{t-1} = \sum_{k=1}^{k} N_k^{t-1}, \pi^{t-1} = \frac{N_k^{t-1}}{N^{t-1}},\\
&{\sigma_k^{t-1}}^2 = \frac{1}{N_k^{t-1}}\sum_{j=1}^{t-1}\sum_{i=1}^{d} z_{ik}^j(x_i^j - \mathbf{u}_i^T\mathbf{v}^j)^2,
\end{aligned} \tag{7}$$

and C is a constant number. So,

$$\begin{aligned}
&\sum_{j=1}^{t} \ln p(\mathbf{x}^j, \mathbf{z}^j|\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{v}^j, \mathbf{U}^j),\\
&= \ln p(\mathbf{x}^t, \mathbf{z}^t|\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{v}^t, \mathbf{U}^t) - \mathcal{R}_F^t(\mathbf{\Pi}, \mathbf{\Sigma}) + C\\
&= -{\mathcal{L}'}^t(\mathbf{\Pi}, \mathbf{\Sigma}) + C.
\end{aligned} \tag{8}$$

Meanwhile, for $\mathbf{U}$, it holds that

$$\begin{aligned}
&\ln p(\mathbf{x}^j, \mathbf{z}^j|\mathbf{\Pi}^j, \mathbf{\Sigma}^j, \mathbf{v}^j, \mathbf{U})\\
&= \sum_{j=1}^{t-1}\sum_{i=1}^{d}\sum_{k=1}^{K} z_{ik}^j\{\ln \pi_k^j - \ln \sigma_k^j - \frac{(x_i^j - \mathbf{u}_i^T\mathbf{v}^j)^2}{2{\sigma_k^j}^2}\} + C\\
&= -||\mathbf{W}^t \odot (\mathbf{X}^t - \mathbf{U}\mathbf{V}^t)||_F^2 + C,
\end{aligned} \tag{9}$$

where $\mathbf{W}^t = [\mathbf{w}^j]_{j=1}^t$,$\mathbf{V}^t = [\mathbf{v}^j]_{j=1}^t$ and $w_i^j = \sqrt{\sum_{k=1}^K \frac{\gamma_{ik}^t}{2{\sigma_k^t}^2}}$. The solution to E.q.(9) is:

$$\begin{aligned}
\mathbf{u}_i^t &= \left(\sum_{j=1}^{t} {w_i^j}^2\mathbf{v}^j{\mathbf{v}^j}^T\right)^{-1}(\sum_{j=1}^{t} {w_i^j}^2 x_i^j{\mathbf{v}^j}^T)\\
&= \left({\mathbf{A}_i^{t-1}}^{-1} + {w_i^j}^2\mathbf{v}^j{\mathbf{v}^j}^T\right)^{-1}(\mathbf{b}_i^{t-1} + {w_i^j}^2 x_i^j{\mathbf{v}^j}^T),
\end{aligned} \tag{10}$$

where ${\mathbf{A}_i^{t-1}}^{-1} = \sum_{j=1}^{t-1} {w_i^j}^2\mathbf{v}^j{\mathbf{v}^j}^T$, $\mathbf{b}_i^{t-1} = \sum_{j=1}^{t-1} {w_i^j}^2 x_i^j{\mathbf{v}^j}^T$, which are also the solution of :

$$ {\mathcal{L}'}^t(\mathbf{U}) = ||\mathbf{w}^t \odot (\mathbf{x}^t - \mathbf{U}\mathbf{v}^t)||_F^2 + \mathcal{R}_b^t(\mathbf{U}). \tag{11}$$

We then prove (5). Firstly we need to make some reasonable and simple bound assumptions for $\forall i, j$:

$$|x_i^j| \le M$$

$$0 < c \le \frac{\sum_{i=1}^d \gamma_{ik}^j}{d} \le C < 1$$

$$0 < \lambda_1 \le w_i^j \le \lambda_2$$

$$0 < \eta_1 \le ||\mathbf{v}^j||_F \le \eta_2.$$

Then, based on Eq. (13) in the paper, we can deduce that when the online EM algorithm converges, we have:

$$\begin{aligned}
&\pi_k^t - \pi_k^{t-1} = \frac{\overline{N}}{N}(\overline{\pi}_k - \pi_k^{t-1});\\
&{\sigma_k^t}^2 - {\sigma_k^{t-1}}^2 = \frac{\overline{N}_k}{N_k}({\overline{\sigma}_k}^2 - {\sigma_k^{t-1}}^2).
\end{aligned} \tag{12}$$

For π_k^t, it is easy to know $|\overline{\pi}_k - \pi_k^{t-1}| \le 1$ and $\frac{\overline{N}}{N} = \frac{1}{t}$, and hence we have:

$$|\pi_k^t - \pi_k^{t-1}| = |\frac{\overline{N}}{N}||\overline{\pi}_k - \pi_k^{t-1}| \le \frac{1}{t}.$$

For ${\sigma_k^t}^2$, it is easy to get $|{\overline{\sigma}_k}^2 - {\sigma_k^{t-1}}^2| \le M^2$ and $\frac{c}{C}\frac{1}{t} \le \frac{\overline{N}_k}{N_k} \le \frac{C}{c}\frac{1}{t}$, and therefore,

$$|{\sigma_k^t}^2 - {\sigma_k^{t-1}}^2| = |\frac{\overline{N}_k}{N_k}||{\overline{\sigma}_k}^2 - {\sigma_k^{t-1}}^2| \le \frac{CM^2}{c}\frac{1}{t}.$$

Thus, we have

$$||\mathbf{\Sigma}^t - \mathbf{\Sigma}^{t-1}||_F \le O(\frac{1}{t}); \quad ||\mathbf{\Pi}^t - \mathbf{\Pi}^{t-1}||_F \le O(\frac{1}{t})$$

For $\mathbf{U}^t$, Since when $\rho = 1$, we have

$$\mathbf{A}_i^{t-1} = \sum_{j=1}^{t} {w_i^j}^2\mathbf{v}^j{\mathbf{v}^j}^T \; \mathbf{b}_i^t = \sum_{j=1}^{t} {w_i^j}^2 x_i^j{\mathbf{v}^j}^T.$$

Therefore, under the bound assumptions, we can get

$$\lambda_1^2\eta_1^2 t \le ||{\mathbf{A}_i^t}^{-1}||_F \le \lambda_2^2\eta_2^2 t,$$

$$\frac{1}{\lambda_2^2\eta_2^2 t} \le ||\mathbf{A}_i^t||_F \le \frac{1}{\lambda_1^2\eta_1^2 t},$$

$$||\mathbf{b}_i^t||_F \le \lambda_2^2\eta_2 M t.$$

By virtue of Eq.(19) in the paper and ${\mathbf{u}_i}^t = \mathbf{A}_i^t\mathbf{b}_i^t$, we can get

$${\mathbf{u}_i}^t - {\mathbf{u}_i}^{t-1} = \mathbf{A}_i^t{w_i^j}^2 x_i^j{\mathbf{v}^j}^T - \mathbf{b}_i^{t-1}\frac{{w_i^t}^2\mathbf{A}_i^{t-1}\mathbf{v}^t{\mathbf{v}^t}^T\mathbf{A}_i^{t-1}}{1 + {w_i^t}^2{\mathbf{v}^t}^T\mathbf{A}_i^{t-1}\mathbf{v}^t}.$$

Calculate F-norm in both sides, we then get:

$$\begin{aligned}||\mathbf{u}_i{}^t-\mathbf{u}_i{}^{t-1}||_F &\le \frac{\lambda_2^2\eta_2 M}{\lambda_1^2\eta_1^2 t}+\lambda_2^2\eta_2 M(t-1)\frac{\lambda_2^2\frac{1}{\lambda_1^4\eta_1^4(t-1)^2}\eta_2^2}{1+\lambda_1^2\eta_1^2\frac{1}{\lambda_2^2\eta_2^2(t-1)}}\\ &=\frac{\lambda_2^2\eta_2 M}{\lambda_1^2\eta_1^2 t}+\frac{\lambda_2^4\eta_2^3 M}{\lambda_1^4\eta_1^4}\frac{1}{t-1+\frac{\lambda_1^2\eta_1^2}{\lambda_2^4\eta_2^4}}\\ &=O(\frac{1}{t})+O(\frac{1}{t-1+\frac{\lambda_1^2\eta_1^2}{\lambda_2^4\eta_2^4}})\\ &=O(\frac{1}{t}).\end{aligned} \tag{13}$$

Thus we have: $||\mathbf{U}^t-\mathbf{U}^{t-1}||_F\le O(\frac{1}{t})$.

The proof is then completed.■

A3: Inference details on MoG Parameter Updating Equations

In this section we introduce how to infer the updating equations on MoG parameters $\{\mathbf{\Pi},\mathbf{\Sigma}\}$ in the M-step of the proposed method (as introduced in Section. 3.3 of the maintext).

After E-step we can get the following objective function:

$$\begin{aligned}\mathcal{L'}^t(\mathbf{\Pi},\mathbf{\Sigma}) &= -\sum_{i=1}^{d}\sum_{k=1}^{K}\gamma_{ik}^t(\ln\pi_k-\ln\sigma_k-\frac{(x_i^t-\mathbf{u}_i^T\mathbf{v})^2}{2\sigma_k^2}))\\ &+(\sum_{k=1}^{K}N_k^{t-1}(\frac{1}{2}\frac{{\sigma_k^{t-1}}^2}{\sigma_k^2}+\ln\sigma_k)-N^{t-1}\sum_{k=1}^{K}\pi_k^{t-1}log\pi_k).\end{aligned} \tag{14}$$

Thus we can deduce that

$$\frac{\partial\mathcal{L'}^t}{\partial\sigma_k}=\sum_{i=1}^{d}\gamma_{ik}^t\{\frac{1}{\sigma_k}-\frac{(x_i^t-\mathbf{u}_i^T\mathbf{v})^2}{2\sigma_k^3}+N_k^{t-1}(-\frac{{\sigma_k^{t-1}}^2}{\sigma_k^3}+\frac{1}{\sigma_k})\}.$$

Let $\frac{\partial\mathcal{L}'}{\partial\sigma_k}=0$, and we have

$$\sigma_k^2=\frac{N_k^{t-1}{\sigma_k^{t-1}}^2+\sum_{i=1}^d\gamma_{ik}^t(x_i^t-\mathbf{u}_i^T\mathbf{v})^2}{N_k^{t-1}+\sum_{i=1}^d\gamma_{ik}^t}. \tag{15}$$

For π_k, note that there is a supplemental constraint $\sum_{k=1}^K\pi_k=1$. We can calculate the derivative on the corresponding Lagrange function as:

$$\frac{\partial\{\mathcal{L}'+\lambda(\sum_{k=1}^K\pi_k-1)\}}{\partial\pi_k}=-\sum_{i=1}^{d}\gamma_{ik}^t\frac{1}{\pi_k}-N_k^{t-1}\frac{1}{\pi_k}+\lambda=0,$$

and then we have

$$N_k^{t-1}+\sum_{i=1}^{d}\gamma_{ik}^t-\lambda\pi_k=0.$$

By summing up k and using $\sum_{k=1}^K\pi_k=1$, we get

$$\lambda=\sum_{k=1}^{K}(N_k^{t-1}+\sum_{i=1}^{d}\gamma_{ik}^t).$$

Thus

$$\pi_k=\frac{N_k^{t-1}+\sum_{i=1}^d\gamma_{ik}^t}{\sum_{k=1}^K(N_k^{t-1}+\sum_{i=1}^d\gamma_{ik}^t)}. \tag{16}$$

TABLE 1: The settings of the rank

Video	$airport$	$bootstrap$	$shoppingmall$	$lobby$	$escalator$
rank	2	2	4	5	2
Video	$curtain$	$campus$	$watersurface$	$fountain$	
rank	6	4	2	4	

If we set

$$\begin{aligned}&\overline{N}=d;\ \overline{N}_k=\sum_{i}^{d}\gamma_{ik}^t;\ \overline{\pi}_k=\frac{\overline{N}_k}{\overline{N}};\\ &\overline{\sigma}_k{}^2=\frac{1}{\overline{N}_k}\sum_{i=1}^{d}\gamma_{ik}^t(x_i^t-(\mathbf{u}_i)^T\mathbf{v})^2);\\ &N=N^{t-1}+\overline{N};N_k=N_k^{t-1}+\overline{N}_k,\end{aligned} \tag{17}$$

Eq. (15) and (16) can then be rewritten as the following forms for $\mathbf{\Pi}$ and $\mathbf{\Sigma}$:

$$\begin{aligned}&\pi_k=\pi_k^{t-1}-\frac{\overline{N}}{N}(\pi_k^{t-1}-\overline{\pi}_k),\\ &\sigma_k{}^2={\sigma_k^{t-1}}^2-\frac{\overline{N}_k}{N_k}({\sigma_k^{t-1}}^2-\overline{\sigma}_k{}^2),\end{aligned} \tag{18}$$

which is the closed-form updating equations for the parameters.

Eq (18) can also be understood in the Robbins-Monro algorithm framework [1], which is an effective methodology in solving the MLE problem on the sequential data.

A4. Complexity Analysis of the OMoGMF Algorithm

All variables involved in the OMoGMF algorithm are updated in closed-form (i.e., Eq. (12), (14), (17), (21) in the main text), and thus we can easily evaluate that the complexity of our method is $O(I(d_\Omega(k+r^2)+r^3))$, where r is the subspace rank, I is the iteration number of the algorithm, d_Ω is the number of pixels in the video frame, and k is the number of noise components. The complexity of the proposed OMoGMF algorithm is thus linearly increasing with d_Ω, k, and I, and three order increasing with r (led by the matrix inverse calculation involved in Eq. (17) in the main text).

A5. More parameters Setting details

A5.1. On Experiment 4.1

We compare OMoGMF and OMoGMF+TV with other nine competing methods, including RPCA, GODEC, RegL1, PRMF, OPRMF, GRASTA,GOSUS and DECOLOR. For RPCA, RegL1, PRMF, OPRMF, DECOLOR and GRASTA, we use the default parameter settings in the original codes. For GODEC, we set the sparse parameter by using the result of RPCA. For GOSUS, we set λ using cross-validation and set others by default settings. The settings of the rank for the methods is shown in Table 1. For OMoGMF and OMoGMF+TV, we set $N^{t-1}=50d$, $\rho=0.98$ and the number of the Gaussians $K=3$. Moreover, the impact of N^{t-1} on MoG parameters is shown in Fig. 1, and we can find that too large or too small N^{t-1} value are not preferred. SPecifically, if it is set too large, the MoG parameters are varied very slowly and are hard to reflect the realtime noise variation of online videos. However, if it is set too small, the parameters are too sensitive to single frame change, and

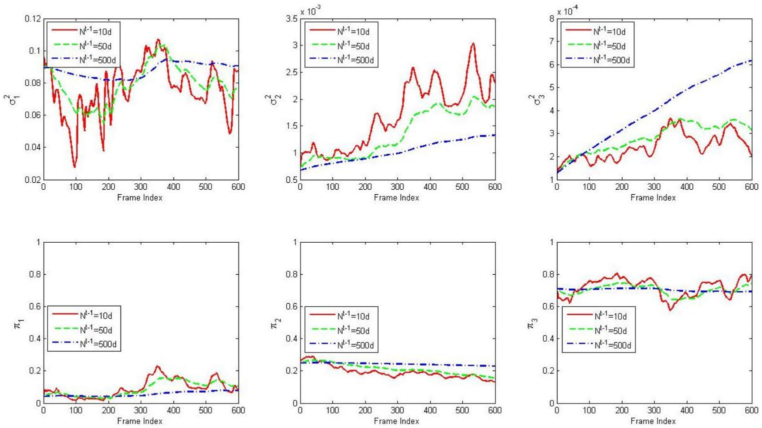

Fig. 1: Tendency curves of MoG parameters with respect to different N_k^{t-1} in $airport$ sequence.

make the method unstable to a continuous change on video foregrounds. We thus more prefer a moderate setting for this parameter, as clearly depicted in the figure. Besides, the threshold number for F-measure is set be the optimal value for all methods, which maximizes the F-measure. We also give the subsampling experiment results of all 9 videos in Table 2 (i.e., the detailed version of Table 5 in the maintext).

A5.2. On Experiment 4.2: Synthetic data

Since RASL and t-GRASTA are not work well on the original frames and these two methods must choose a canonical frame which is smaller than original frames. In order to fairly compare t-OMoGMF with them, we choose a $0.76m \times 0.76n$ canonical frame, where m and n are the length and width of original frames, respectively. Meanwhile, we also use the same size of canonical frame on ground truth and the respective image transformation obtained by these three methods to compute the F-measure. Besides, we use it-OMoGMF as a warm-start way to t-OMoGMF and set the number of the Gaussians as $K = 2$.

A5.3. On Experiment 4.2: Real data

For RASL and t-GRASTA, we choose a $0.7m \times 0.7n$ canonical frame on the real data experiments. We randomly choose 50 frames and use it-OMoGMF to warm start the model, while set the number of the Gaussians as $K = 2$.

References

[1] H. Robbins and S. Monro, “A stochastic approximation method,” *The annals of mathematical statistics*, pp. 400–407, 1951.

TABLE 2: F-measure and FPS of OMoGMF and GRASTA under different sub-sampling rates on 9 vidoes, each with 1000 frames, in Li dataset.

Sub-Sampling rate			1%		10%		30%		50%		100%	
Dataset	frame size	method	F-M	FPS	F-M	FPS	F-M	FPS	F-M	FPS	F-M	FPS
$airport$	144× 176	OMoGMF	0.7131	263.6	0.7230	181.6	0.7238	115.7	0.7242	91.7	0.7241	63.0
		GRASTA	0.6312	246.7	0.6403	209.6	0.6325	167.6	0.6256	141.3	0.6194	123.9
$bootstap$	120× 160	OMoGMF	0.6196	334.5	0.6204	273.4	0.6204	178.7	0.6214	144.4	0.6199	104.3
		GRASTA	0.5462	319.3	0.5757	279.4	0.5812	224.2	0.5832	191.1	0.5788	177.9
$shoppingmall$	256× 320	OMoGMF	0.6942	104.7	0.6946	40.2	0.6947	16.1	0.6948	10.2	0.6950	5.2
		GRASTA	0.6535	65.5	0.7100	56.7	0.7143	44.2	0.7151	37.1	0.7146	28.7
$lobby$	128× 160	OMoGMF	0.7699	276.7	0.7716	173.2	0.7721	112.5	0.7714	68.1	0.7721	41.1
		GRASTA	0.2389	246.2	0.5251	218.9	0.6290	174.7	0.6295	152.3	0.6101	132.9
$escalator$	130× 160	OMoGMF	0.6101	332.0	0.6119	265.9	0.6126	172.3	0.6128	137.5	0.6120	99.6
		GRASTA	0.4581	303.2	0.5897	264.6	0.5870	212.8	0.5821	180.2	0.5732	166.9
$curtain$	128× 160	OMoGMF	0.8558	266.4	0.8669	159.8	0.8647	63.9	0.8657	44.1	0.8657	24.7
		GRASTA	0.7828	222.9	0.7908	197.3	0.7442	157.6	0.7143	133.7	0.6830	113.5
$campus$	128× 160	OMoGMF	0.4363	270.3	0.4465	166.6	0.4474	75.4	0.4470	52.6	0.4475	30.9
		GRASTA	0.3795	233.6	0.4108	203.2	0.4287	163.0	0.4357	139.5	0.4458	118.4
$watersurface$	128× 160	OMoGMF	0.8716	331.3	0.8728	267.5	0.8744	174.6	0.8744	138.41	0.8744	100.2
		GRASTA	0.8527	301.2	0.8564	266.3	0.8462	215.2	0.8264	185.6	0.7589	168.8
$fountain$	128× 160	OMoGMF	0.7136	281.1	0.7196	176.8	0.7197	118.7	0.7196	88.2	0.7196	49.5
		GRASTA	0.6013	257.4	0.6881	229.3	0.7016	180.5	0.6999	150.5	0.6923	127.5
$Average$	–	OMoGMF	**0.6982**	**273.4**	**0.7030**	189.4	**0.7033**	114.2	**0.7035**	86.1	**0.7034**	57.6
		GRASTA	0.5716	244.0	0.6430	**213.9**	0.6516	**171.1**	0.6458	**145.7**	0.6307	**128.7**