

# Temporal Compressive Video Reconstruction Using Gaussian Scale Mixture Model

Xiao-hai HE<sup>1,\*</sup> and Mao-jiao WANG<sup>1,2</sup>

<sup>1</sup>Image Information Institute, College of Electronics and Information Engineering,  
Sichuan University, No. 24 South Section 1, Yihuan Road, Chengdu, 610065, China

<sup>2</sup>Dean's office, Panzhihua University, Panzhihua, 617000, China

\*Corresponding author: hxx@scu.edu.cn

**Keywords:** Compressive sensing, Video reconstruction, Temporal compressive measurements, Gaussian scale mixture.

**Abstract.** Compressive sensing has been used to acquire the information in high-frame-rate video using low-frame-rate compressive measurements. Under the framework of coded aperture compressive temporal imaging, we propose a video reconstruction algorithm using Gaussian scale mixture model from temporal compressive measurements. Experimental results demonstrate that our proposed algorithm outperforms state-of-the-art algorithms in both peak signal-to-noise ratio and visual quality.

## Introduction

Measuring high-speed video is a challenge to camera design. Recently, compressive sensing (CS) [1-3] has been used to acquire the information in high-frame-rate video using low-frame-rate compressive measurements. A number of inverse algorithms have been proposed for compressive video reconstruction lately. Park and Wakin [4] described a coarse-to-fine inverse algorithm which interchanged between temporal motion estimation and spatial video reconstruction in the wavelet domain. Yang et al. [5,6] utilized a Gaussian mixture model to represent spatiotemporal video patches, and the video reconstruction can be efficiently calculated based on analytic expressions. In this paper, we extend Gaussian scale mixture (GSM) [7] model to compressive video reconstruction. While doing CS video reconstruction, the video's pixel volume is partitioned into a set of spatiotemporal patches. Since the dimension of each video patch is much larger than that of the corresponding compressive measurement, all pixels within a patch are assumed to be drawn from a GSM model.

## Temporal Compressive Video Measurements

Because of modulating high-speed video at low power and cost, the coded aperture compressive temporal imaging (CACTI) [8-10] is utilized to obtain temporal compressive video measurements. Let  $\rho_x$  and  $\rho_y$  denote the number of pixels in the horizontal and vertical axes of the video frame.  $F$  is a parameter that may be set depending on what is happening in the scene.  $\mathbf{X} \in \mathbb{R}^{\rho_x \times \rho_y \times F}$  describes the discretized reconstructed video, and  $\mathbf{Y} \in \mathbb{R}^{\rho_x \times \rho_y}$  represent the temporal compressive video measurements. The CACTI camera sums  $F$  coded frames to a two-dimensional (2D) temporal compressive measurement. More formally

$$\mathbf{Y}_{i,j} = [\mathbf{H}_{i,j,1} \mathbf{H}_{i,j,2} \cdots \mathbf{H}_{i,j,\kappa}] \circ [\mathbf{X}_{i,j,1} \mathbf{X}_{i,j,2} \cdots \mathbf{X}_{i,j,\kappa}]^T. \quad (1)$$

where  $\mathbf{Y}_{i,j}$  is the component  $(i, j)$  of measurements  $\mathbf{Y}$ ,  $\mathbf{H}$  denotes the binary mask with components  $\mathbf{H}_{i,j,k} \in \{0,1\}$ , which is designed by randomly choosing 1/0 values at each pixel, with a 0.5 probability of 1.  $\mathbf{X}_{i,j,k}$  is connected with the same pixels  $(i, j)$ , and temporal bin  $k$ . The symbol  $\circ$  represents Hadamard product, and  $^T$  denotes transpose of a matrix. Thus, each pixel in  $\mathbf{Y}$  is comprised of a weighted sum of the corresponding pixel values in the  $F$  frames of  $\mathbf{X}$ , where the weights are binary. In this paper, our objective is to recover  $F$  coded frames from each compressive measurement.

### Video Reconstruction Using GSM Model

We divide  $\mathbf{Y}$  into a total of  $M$  2D patches, where each square patch constitutes  $P \times P$  pixels. Let  $y_m \in \mathbb{R}^{P^2}$  be the vectorized  $m^{th}$  patch.  $x_m \in \mathbb{R}^{P^2 F}$  represents the corresponding vectorized spatiotemporal patches to be recovered. Thus, each compressive measurement can be expressed as

$$y_m = \Phi_m x_m + \xi_m \quad \forall m = 1, 2, \dots, M. \quad (2)$$

where  $\Phi_m \in \{0,1\}^{P^2 \times P^2 F}$  is measuring matrix projecting  $x_m$  onto  $y_m$ , and  $\xi_m \in \mathbb{R}^{P^2}$  denotes zero-mean Gaussian measurement noise with variance  $\sigma_\eta^2$ . For each  $m$ , our goal is to recover  $x_m$  from the patch  $y_m$ . To solve this problem, we assume that all pixels within  $x_m$  are drawn from a GSM model.

### GSM Model

A vector  $x_m \in \mathbb{R}^N$  ( $N = P^2 F$ ) is a GSM [7] model if and only if it can be expressed in the form  $x_m = \omega^{1/2} G$ , where  $\omega$  is an independent positive scalar random variable and  $G$  is a Gaussian random vector, with covariance matrix  $\Sigma = \text{diag}(\delta_1^2, \delta_2^2, \dots, \delta_N^2)$ . The probability density function (pdf) of  $x_m$  conditioned on random variable  $\omega$  and covariance matrix  $\Sigma$  is denoted by

$$p(x_m | \omega, \Sigma) = (2\pi)^{-N/2} |\omega \Sigma|^{-1/2} \exp \left[ -\frac{1}{2} x_m^T (\omega \Sigma)^{-1} x_m \right]. \quad (3)$$

where  $|\cdot|$  represents determinant of a matrix. From Eq. 3, we can obtain the maximum likelihood estimate of random variable  $\omega$ :

$$\omega(x_m) = N^{-1} x_m^T \Sigma^{-1} x_m. \quad (4)$$

### Reconstruction Algorithm Using GSM Model

From Eq. 2, we have the Gaussian likelihood model of compressive measurement  $y_m$ . More formally

$$p(y_m|x_m, \sigma_\eta^2) = (2\pi\sigma_\eta^2)^{-p^2/2} \exp\left(-\frac{1}{2\sigma_\eta^2} \|y_m - \Phi_m x_m\|_2^2\right). \quad (5)$$

If the noise variance  $\sigma_\eta^2$ , the value of  $\omega$  and covariance matrix  $\Sigma$  have been estimated, given the compressive measurement  $y_m$  and measuring matrix  $\Phi_m$ , the posterior pdf of  $x_m$  is given by

$$p(x_m|y_m, \omega, \Sigma, \sigma_\eta^2) = \frac{p(x_m|\omega, \Sigma) p(y_m|x_m, \sigma_\eta^2)}{p(y_m|\omega, \Sigma, \sigma_\eta^2)}. \quad (6)$$

Substituting Eq. 3 and Eq. 5 into Eq. 6, we can obtain a multivariate Gaussian distribution, with covariance  $\Theta$  and mean  $\mu$  given by

$$\Theta = (\sigma_\eta^{-2} \Phi_m^T \Phi_m + \Gamma)^{-1}, \quad \mu = \sigma_\eta^{-2} \Gamma \Phi_m^T y_m. \quad (7)$$

where  $\Gamma = \text{diag}\left[(\omega\delta_1^2)^{-1}, (\omega\delta_2^2)^{-1}, \dots, (\omega\delta_N^2)^{-1}\right]$ . Thus, the reconstructed vector  $x_m$  is chosen to be equal to  $\mu$ . Therefore, the problem only needs to estimate the unknown model parameters  $\omega, \Sigma$  and  $\sigma_\eta^2$ . Noting that the covariance  $\Theta$  can be rewritten as its equivalent form

$$\omega^{-1} \Theta = (\omega\sigma_\eta^{-2} \Phi_m^T \Phi_m + \Sigma^{-1})^{-1}. \quad (8)$$

So we can estimate the unknown model parameters  $\sigma_\eta^2$  and  $\{\delta_i^{-2}\}_{i=1}^N$  by maximizing the following marginal log-likelihood function:

$$\begin{aligned} \Upsilon(\sigma_\eta^2, \{\delta_i^{-2}\}_{i=1}^N) &= \log \left[ p(y_m|\omega, \sigma_\eta^2, \{\delta_i^{-2}\}_{i=1}^N) \right] \\ &= -\frac{1}{2} \left[ p^2 \log(2\pi) + \log|\Pi| + y_m^T \Pi^{-1} y_m \right] \end{aligned} \quad (9)$$

where  $\Pi = \omega^{-1} \sigma_\eta^2 \mathbf{I} + \Phi_m \Sigma \Phi_m^T$ .

Next, we can decompose matrix  $\Pi$  as follows:

$$\Pi = \omega^{-1} \sigma_\eta^2 \mathbf{I} + \sum_{v=1, v \neq i}^N \delta_v^2 \Phi_{m_v} \Phi_{m_v}^T + \delta_i^2 \Phi_{m_i} \Phi_{m_i}^T = \Pi_{-i} + \delta_i^2 \Phi_{m_i} \Phi_{m_i}^T. \quad (10)$$

where  $\Pi_{-i}$  is  $\Pi$  with the contribution of  $i^{th}$  basis vector removed, and  $\Phi_{m_i}$  denotes the  $i^{th}$  column of  $\Phi_m$ . Using matrix determinant lemma and Sherman-Morrison-Woodbury identity, we have

$$|\Pi| = |\Pi_{-i}| \left| 1 + \delta_i^2 \Phi_{m_i}^T \Pi_{-i}^{-1} \Phi_{m_i} \right|, \quad \Pi^{-1} = \Pi_{-i}^{-1} - \frac{\Pi_{-i}^{-1} \Phi_{m_i} \Phi_{m_i}^T \Pi_{-i}^{-1}}{\delta_i^{-2} + \Phi_{m_i}^T \Pi_{-i}^{-1} \Phi_{m_i}}. \quad (11)$$

The fast update algorithm [11, 12] is utilized to increase monotonically the log-likelihood function, which can be rewritten as

$$\begin{aligned} \Upsilon\left(\sigma_{\eta}^2, \left\{\delta_i^{-2}\right\}_{i=1}^N\right) &= -\frac{1}{2}\left[N \log (2 \pi)+\log \left|\Pi_{-i}\right|+y_m^T \Pi_{-i}^{-1} y_m\right] \\ &+ \frac{1}{2}\left[\log \left(\delta_i^{-2}\right)-\log \left(\delta_i^{-2}+S_i\right)+\frac{Q_i^2}{\delta_i^{-2}+S_i}\right]=\Upsilon\left(\sigma_{\eta}^2, \left\{\delta_v^{-2}\right\}_{v=1, v \neq i}^N\right)+\gamma\left(\delta_i^{-2}\right) \end{aligned} \quad (12)$$

where  $S_i = \Phi_{m_i}^T \Pi_{-i}^{-1} \Phi_{m_i}$  and  $Q_i = \Phi_{m_i}^T \Pi_{-i}^{-1} y_m$ .

It can be seen that  $\Upsilon\left(\sigma_{\eta}^2, \left\{\delta_i^{-2}\right\}_{i=1}^N\right)$  has a unique maximum with respect to  $\delta_i^{-2}$ :

$$\delta_i^{-2} = \begin{cases} \frac{S_i^2}{Q_i^2 - S_i}, & \text{if } Q_i^2 > S_i, \\ \infty, & \text{if } Q_i^2 \leq S_i. \end{cases} \quad (13)$$

The GSM-based video reconstruction algorithm is summarized as follows:

Step 1: set  $\sigma_{\eta}^2 = 10^{-3} \cdot \text{variance}(y_m)$ .

Step 2: initialize a single basis vector  $\Phi_{m_{i_1}}$  such that

$$i_1 = \arg \max_{i=1,2,\dots,N} \frac{\left\|\Phi_{m_i}\right\|^2}{\left(\left\|\Phi_{m_i}^T y_m\right\|^2 / \left\|\Phi_{m_i}\right\|^2\right) - \sigma_{\eta}^2}, \text{ and set } \Lambda = \left\{i_1\right\}.$$

Step 3: set  $\delta_{i_1}^{-2} = \frac{\left\|\Phi_{m_{i_1}}\right\|^2}{\left(\left\|\Phi_{m_{i_1}}^T y_m\right\|^2 / \left\|\Phi_{m_{i_1}}\right\|^2\right) - \sigma_{\eta}^2}$ , and all other  $\left\{\delta_i^{-2}\right\}_{i=1, i \neq i_1}^N$  are set to infinity.

Step 4: using Eq. 7 and Eq. 4 respectively, compute  $\Theta$ ,  $\mu$  and  $\omega$ .

Step 5: compute  $\theta_i = Q_i^2 - S_i$ .

Step 6: if  $\theta_i > 0$  and  $\delta_i^{-2} < \infty$ , re-estimate  $\delta_i^{-2}$ .

Step 7: if  $\theta_i > 0$  and  $\delta_i^{-2} = \infty$ , add  $\Phi_{m_i}$  in the set  $\Lambda$  and update  $\delta_i^{-2}$ .

Step 8: if  $\theta_i \leq 0$  and  $\delta_i^{-2} < \infty$ , delete  $\Phi_{m_i}$  from the set  $\Lambda$  and set  $\delta_i^{-2} = \infty$ .

Step 9: update  $\Theta$ ,  $\mu$  and  $\omega$ .

Step10: update  $\sigma_{\eta}^2 = \frac{\left\|y_m - \Phi_m \mu\right\|^2}{N - \text{card}(\Lambda) + \sum_{\tau \in \Lambda} \omega^{-1} \delta_{\tau}^{-2} \Theta_{\tau \tau}}$ , where  $\text{card}(\cdot)$  is the

cardinality of a set.

Step11: update  $\Gamma$  according to the scaling  $\omega \delta_i^2$ .

Step12: if converged terminate, otherwise go to step 5.

## Experimental Results

The proposed GSM-based inverse algorithm reconstructs 8 temporal frames per single compressive video measurement, which is divided into a set of overlapping patches of size  $8 \times 8$ . We use PSNR to assess the quality of every reconstructed frame, and finally the average value over all reconstructed video frames is taken as the evaluation criterion for video reconstruction. The experiment is done on an Intel i5-4590 CPU running at 3.30 GHz with 4GB RAM.

Fig. 1 shows reconstructed frames 89 and 90 using Yang et al.'s method [5] and our

proposed GSM-based method from the last compressive measurements of traffic dataset. When the proposed GSM-based method is used, the reconstructed frames not only have higher PSNR than that of Yang et al.'s method, but also have better visual quality. Furthermore, we also have that the average PSNR over all reconstructed frames of Yang et al.'s method is 21.93dB, while that of our proposed algorithm is 22.41dB.

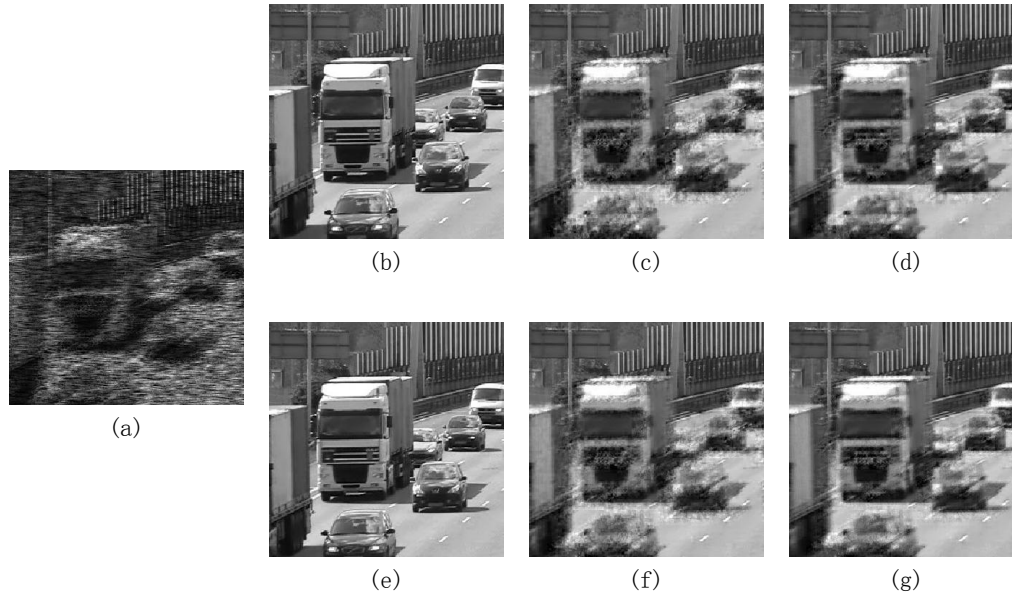


Fig. 1. Reconstruction results of the simulated traffic measurements. (a) Raw measurement. (b) Original frame 89. (c) Reconstructed frame 89 by Yang et al.'s method, PSNR = 21.83 dB. (d) Reconstructed frame 89 by the proposed GSM-based method, PSNR = 22.57 dB. (e) Original frame 90. (f) Reconstructed frame 90 by Yang et al.'s method, PSNR = 22.27 dB. (g) Reconstructed frame 90 by the proposed GSM-based method, PSNR = 23.08 dB.

## Conclusions

In this paper, we have proposed the video reconstruction algorithm using GSM model from temporal compressive measurements. The GSM model is used to represent spatiotemporal patches, and a fast update algorithm is utilized to increase monotonically the log-likelihood function. Experimental results demonstrate that our proposed algorithm outperforms state-of-the-art algorithms in both peak signal-to-noise ratio and visual quality.

## Acknowledgements

The authors would like to thank the anonymous reviewers for their thorough and valuable comments and suggestions. This work was supported by the National Natural Science Foundations of China (No.61471248), the Project of application foundation in Sichuan province, China (No.2015JY0189), and the Innovation project on the education for graduate students of Sichuan province, China (2014-Edu.-034).

## References

- [1] E. J. Candès, J. K. Romberg, and T. Tao, IEEE Trans. Inf. Theory, 52(2), (2006) 489-509.
- [2] D. L. Donoho, IEEE Trans. Inf. Theory, 52(4), (2006) 1289-1306.
- [3] E. J. Candès, and T. Tao, IEEE Trans. Inf. Theory, 52(12), (2006) 5406-5425.

- [4] J. Y. Park, and M. B. Wakin, *J. Electron. Imag.*, 22(2), (2013) 57-61.
- [5] J. Yang, X. Yuan, X. Liao, et al., *IEEE Trans. Image Proces.*, 23(11), (2014) 4863-4878.
- [6] J. Yang, X. Yuan, X. Liao, et al., Gaussian Mixture Model for Video Compressive Sensing, *IEEE International Conference on Image Proces.*, Melbourne, pp. 19-23, Sep. 2013.
- [7] D. F. Andrews and C. L. Mallows, *J. Roy. Statist. Soc.*, 36(1), (1974) 99-102.
- [8] P. Llull, X. Liao, X. Yuan, et al., *Optics Express*, 21(9), (2013) 10526-10545.
- [9] P. Llull, X. Liao, X. Yuan, et al., Compressive sensing for video using a passive coding element, *Proceeding of Computational Optical Sensing and Imaging*, pp. 1-3, Jun. 2013.
- [10] Y. Hitomi, J. Gu, M. Gupta, et al., Video from a single coded exposure photograph using a learned over-complete dictionary, *IEEE International Conference on Computer Vision*, Barcelona, pp. 287-294, Nov. 2011.
- [11] M. E. Tipping and A. C. Faul, Fast marginal likelihood maximization for sparse Bayesian models, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Florida, Jan. 2003.
- [12] M. Saarela, T. Elomaa and K. Ruohonen, *Advances in Machine Learning I. Studies in Computational Intelligence 262*, Springer-Verlag, Berlin, 2010, pp.227-246.