Data Compression for optical movie data of the Tomo-e Gozen

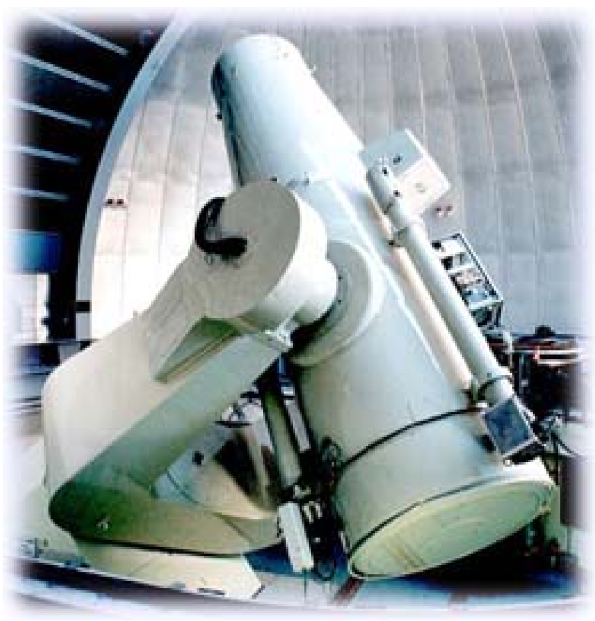
Mikio Morii[1], Shiro Ikeda[1], Ryou Ohsawa[2], and Shigeyuki Sako[2]

[1]The Institute of Statistical Mathematics, Tokyo, Japan

[2]Institute of Astronomy, The University of Tokyo, Tokyo, Japan

The Tomo-e Gozen is a wide-field camera on 1-m Schmidt telescope on Kiso Observatory at Japan. By using the camera, we plan to monitor the wide area of the sky with high speed of 2 Hz, to search optical transients. Full-scale observation will start on 2018. The camera consists of 84 CMOS chips and cover the fields of view with 20 deg^2. The data obtained by one night observation will become about 30 TB. Then, data compression is necessary. We have applied a matrix factorization method for the movie data and showed that the method can compress the data into one-tenth of original size, without losing scientifically important information. We demonstrated that the method can process the data with sufficiently high speed.

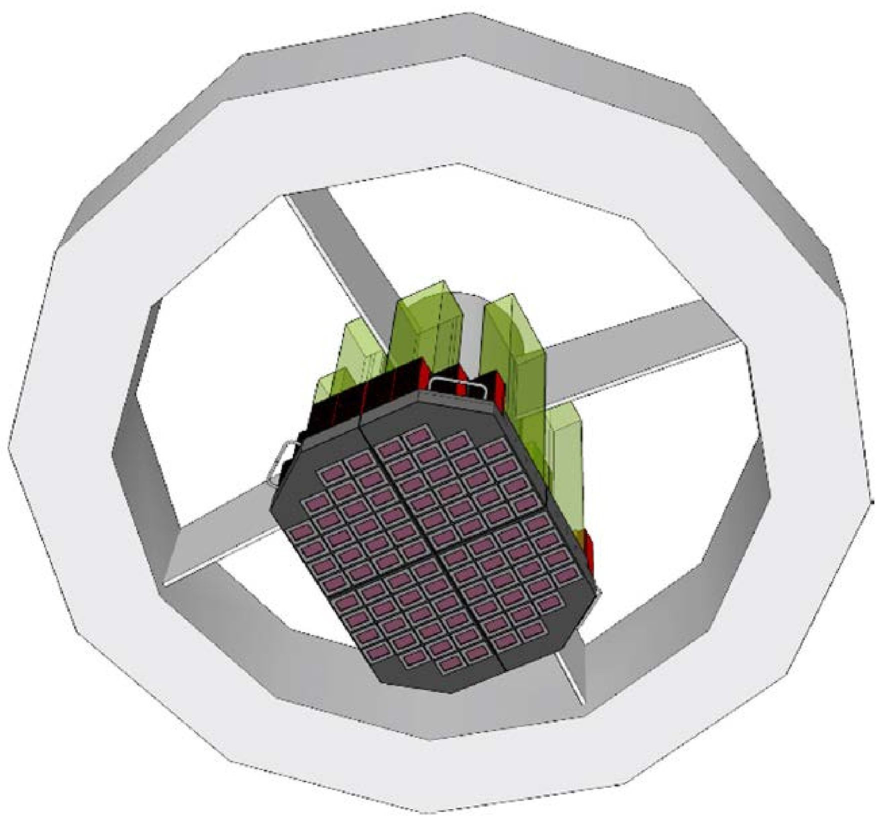
Kiso Observatory



105cm Schmidt Telescope

Kiso Observatory is located at a middle area of Japan.

The Tomo-e Gozen



The Tomo-e Gozen is a wide-field CMOS camera on the Kiso 1.0-m Schmidt telescope under development. It will be completed in 2018. It is optimized for movie observations with sub-second to seconds time-resolutions. It will take consecutive frames with a field-of-view of 20 deg^2 at 2 frames per second (fps) by 84 chips of 2k x 1k CMOS sensors. Now, a prototype model of the Tomo-e Gozen (Tomo-e PM) with 8 CMOS chips is operated.

The primary objective of observations with the Tomo-e Gozen is to catch rare and fast transient phenomena with a time duration shorter than 10 seconds, such as optical counterparts of fast radio bursts, gamma-ray bursts, binary neutron star mergers, etc. The wide field movie observation with the Tomo-e Gozen is also ideal for exploring fast moving objects, including near earth objects and space debris.

The full-scale Tomo-e Gozen will produce a huge amount of data of 30 TB per night. Therefore, the data compression is necessary. Details are shown in [1, 2].

Method

We choose to use low-rank matrix approximation to compress the data size without losing transient events. It was originally proposed by [3]. We use GoDec [4] instead of [3]. We applied it for a movie data of Tomo-e PM with a format of FITS/cube. We re-arrange the pixel values contained in a movie data set (n_x x n_y pixels x n_f frames) into a matrix with n_x n_y rows and n_f columns. Here, n_x =1136, n_y = 2008, and n_f = 400. The number of pixels can be reduced to

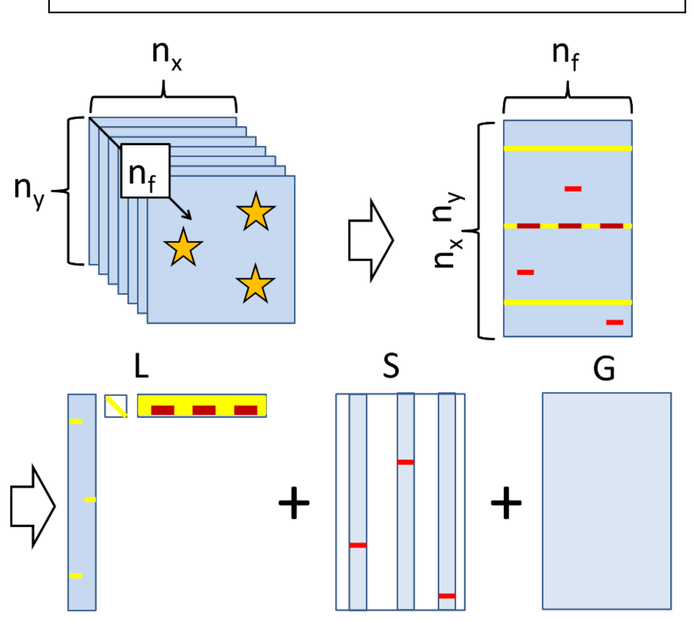
$$n_{\mathrm{red}} = r n_x n_y + r n_f + r + n_{\mathrm{f,sp}} n_x n_y$$

where r is a rank of the low-rank matrix L, and $n_{f,sp}$ is the number of frames of the sparse matrix S, which contains transient events. Typically, r is about 10 and $n_{f,sp}$ is smaller than 10. Therefore, the reduction factor of the data is about 1/10.

GoDec decompositon

$$X = L + S + G$$

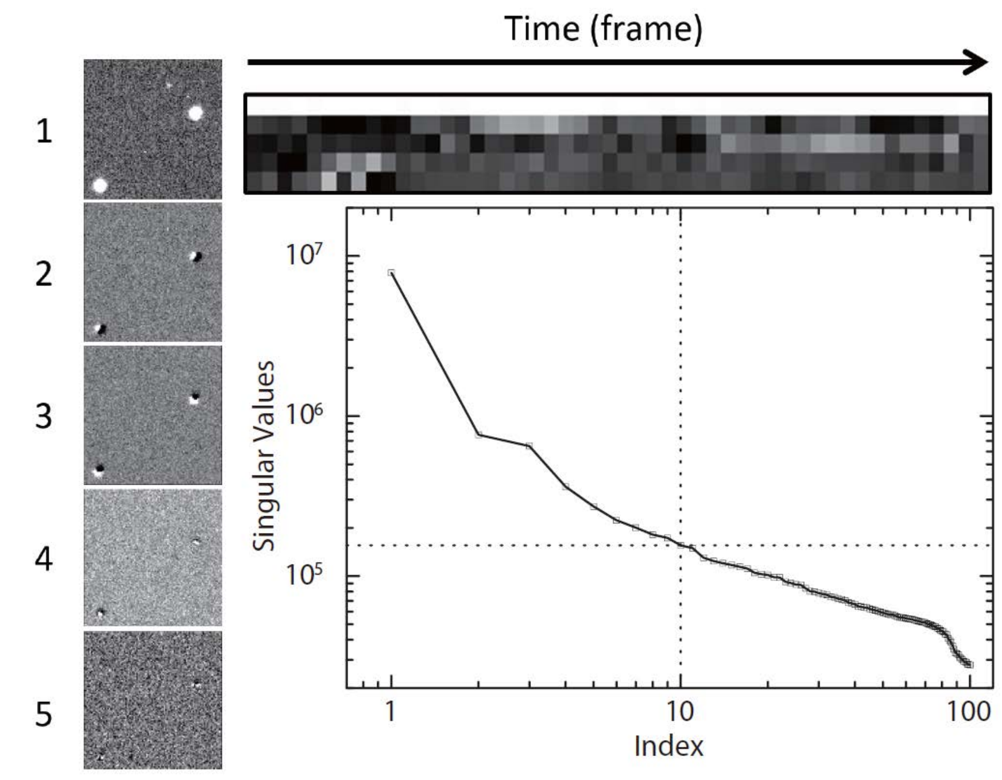
$$\min_{L,S} \quad \|X - L - S\|_F^2 \quad s.t. \quad \mathrm{rank}\,(L) \le r, \quad \mathrm{card}\,(S) \le k.$$



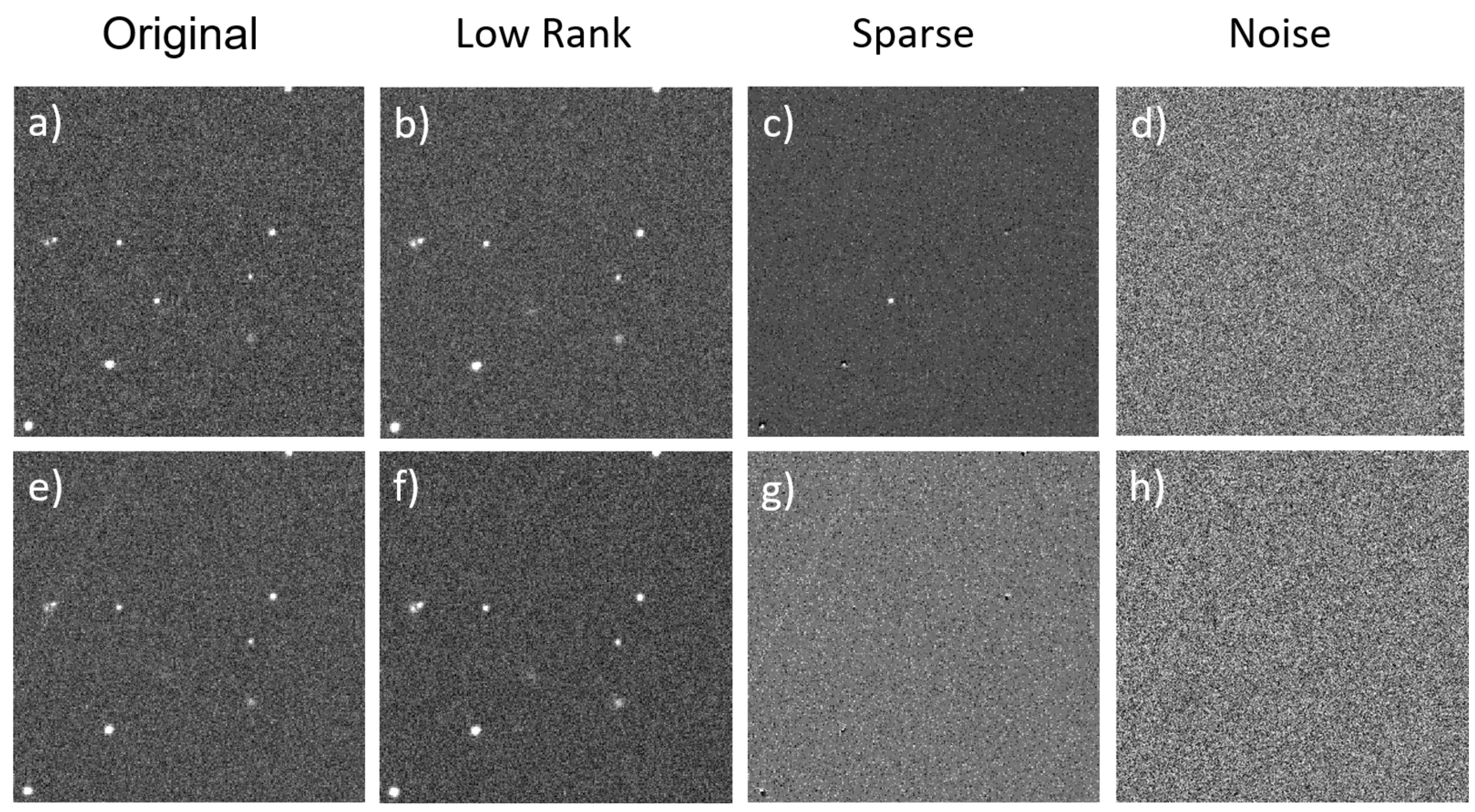
Schematic illustration of data conversion from cube data to a matrix (M), and the matrix decomposition into a low-rank (L), sparse (S) and noise matrix (G). The low-rank matrix is further decomposed by SVD into L = UDV^T .

Structure of low-rank matrix

The central panel shows the distribution of singular values of the matrix obtained by SVD as a function of the index of the diagonal matrix of SVD. The five images in the left-hand side show stable point sources with the top five singular values, where each image is the decomposed component by SVD. The upper panel shows the time variation of intensity of these components in gray scale. We set the rank of the low-rank matrix to 10.



Example of decomposition images



Original (M), low-rank (L), sparse (S), and noise (G) images are shown in the four columns in this respective order from the left. A transient point source appears near the center of the image at the time-frame of the top row, as spotted in the original image (a) in contrast to (e) taken in a different frame (bottom row), and as clearly visible in the sparse one (c) in contrast to (g). On the other hand, a line, which is a light trail caused by a meteoroid, is seen only in the second time-frame (bottom row), as in the original image (e) and sparse one (g). These transient events are not recognized in the low-rank images (b, f). Noise images (d, h) do not contain any patterns noticeable. These images demonstrate that the decomposition is successful.

Speed of processing

The processing time of the decomposition was 320 s for the movie data with 1136 x 2008 pixels and 400 frames, by setting 10 for the rank of a low-rank matrix, for which we used a budget computer, equipped with a CPU of Intel Xeon Processor E5-1630 (3.70 GHz) consisting of eight cores. The processing time is 1.6 times longer than the duration of observation of 200 s (0.5 s x 400 frames), which is sufficiently fast for daily observations.

Conclusion

We have proposed to use the low-rank matrix approximation with sparse matrix decomposition in order to compress the movie astronomical data without losing transient events. Compared with the conventional low-rank matrix approximation with principal component analysis and SVD, our method has an advantage of preserving the transients in the sparse matrices, which is essential for the transient-search project with the Tomo-e Gozen. Details are shown in [5].

References:

[1] Sako, S., et al., 2016, SPIE, 9908, 99083P
[2] Ohsawa, R., et al., 2016, SPIE, 9913, 991339
[3] Candes, E., et al., 2009 http://www-stat.stanford.edu/ candes/papers/RobustPCA.pdf
[4] Zhou, T. & Tao, D. 2011, 28th ICML, 33
[5] Morii, M., Ikeda, S., Sako, S. & Ohsawa, R., 2017, Astrophysical Journal, 835, 1