Astroinformatics
Proceedings IAU Symposium No. 325, 2016
M. Brescia, S.G. Djorgovski, E. Feigelson,
G. Longo & S. Cavuoti, eds.

doi:10.1017/S1743921316012989

An Effective Method for Modeling Two-dimensional Sky Background of LAMOST

Hasitieer Haerken[1,2], Fuqing Duan[1], Jiannan Zhang[2] and Ping Guo[1]

[1]Image Processing and Pattern Recognition Laboratory, Beijing Normal University, 100875, Beijing, China
email: fqduan@bnu.edu.cn pguo@ieee.org

[2]National Astronomical Observatories & Chinese Academy of Sciences
100012 Beijing, China
email: hastear@bao.ac.cn jnzhang@bao.ac.cn

Abstract. Each CCD of LAMOST accommodates 250 spectra, while about 40 are used to observe sky background during real observations. How to estimate the unknown sky background information hidden in the observed 210 celestial spectra by using the known 40 sky spectra is the problem we solve. In order to model the sky background, usually a pre-observation is performed with all fibers observing sky background. We use the observed 250 skylight spectra as training data, where those observed by the 40 fibers are considered as a base vector set. The Locality-constrained Linear Coding (LLC) technique is utilized to represent the skylight spectra observed by the 210 fibers with the base vector set. We also segment each spectrum into small parts, and establish the local sky background model for each part. Experimental results validate the proposed method, and show the local model is better than the global model.

Keywords. LAMOST, celestial spectra, sky background modeling

1. Introduction

LAMOST (The Large Sky Area Multi-Object Fiber Spectroscopic Telescope), a special quasi-meridian reflecting Schmidt telescope locates in Xinglong Station of National Astronomical Observatory, China, with the main goals of (Su *et al.* 1998): (1) obtaining numerous spectrum data of galaxies and QSOs for the research of contemporary cosmology; (2) observing the spectra of many variable objects repeatedly; (3) identifying a large number of objects found by xray, IR, radio and other surveys. More detailed information about LAMOST has been described elsewhere (Cui *et al.* 2000, Cui 2006, Cui *et al.* 2010, Zhao *et al.* 2012).

Sky has been defined to be all the photons which enter the detector coincident in the measured coordinates, one or two spatial dimensions, time, and wavelength, with that of the desired object, but which did not originate in the object (Wyse *et al.* 1994). For astronomical observations, night sky spectra, in most cases will adversely affect the observation, as the background, which will reduce the spectrum SNR. Therefore, sky subtraction has become the primary problem to solve for the observation(Cuby 1994, Watson *et al.* 1998).

The focal surface of LAMOST accommodates 4000 precisely positioned fibers to observe 4000 objects simultaneously, and there are only a few fibers for skylight observation. To obtain pure object spectra with high quality, it is essential to remove the sky background components from the observed spectra. However, the sky backgrounds randomly change with time and spatial differences, which increase the difficulty of sky subtraction. Therefore, our work is aimed at solving the sky background modeling problem.

2. Modeling Method

2.1. *Spectrum extraction*

In order to carry out two-dimensional modeling, the first thing we need to do is to extract the spectra information from the two-dimensional spectral image, that is, spectrum extraction. There are usually three methods for spectrum extraction: aperture extraction method (Horne 1986, Robertson 1986), profile fitting method (Piskunov *et al.* 2002, Blondin *et al.* 2004), and deconvolution extraction method (Bolton *et al.* 2010, Yu *et al.* 2014, Li *et al.* 2015). Li *et al.* (2011) have compared these methods and consider that with the development of deconvolution method, maybe it will gradually replace the other two methods, as the only reliable extraction spectrum method. However, the current spectrum extraction method widely used for LAMOST is still the aperture extraction method.

2.2. *Modeling*

Although the sky backgrounds randomly change with time and spatial differences, when the observation area is small, it can be considered that within this region the skylight of different position is the same. The focal surface of LAMOST can accommodate 4000 optical fibers for observation, and there are 16 spectrographs with 32 CCD cameras to record both red and blue images of multi fiber spectra (Cui *et al.* 2012). On each CCD image, 250 fiber spectra can be accommodated, during real observations, while about 40 are used to observe sky background. In order to model the sky background, usually a pre-observation is performed with all fibers observing sky background. We use the observed 250 skylight spectra as training data, where those observed by the 40 fibers are considered as a base vector set. We represent the skylight spectra observed by the 210 fibers with the base vector set.

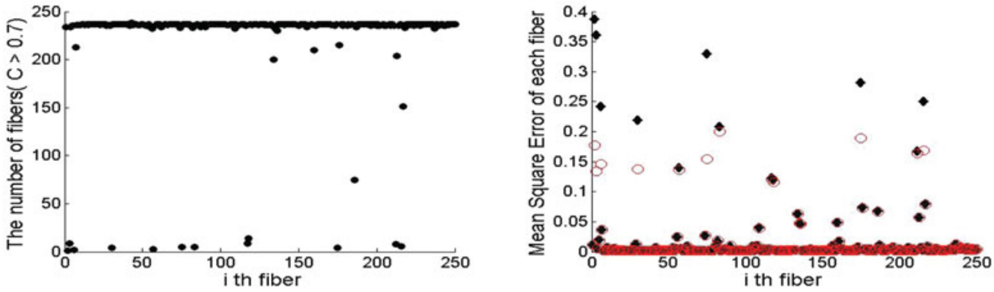
2.3. *Parameter optimization*

We have found through experiments that in the modeling process, if the 40 base vectors are utilized at the same time, the weight of each base vector will become smaller and skylight spectrum fitting effect will not be ideal on the contrary. In order to increase the accuracy, we utilize the most correlated sky spectra of each spectrum to complete modeling, and the essence of our modeling is to find the similarity between the spectra. Both the correlation coefficient, a statistical indicators used to reflect the degree of correlation relationship between variables, and the Euclidean distance can be used for similarity measure.

LLC method, proposed by Wang *et al.* (2010) utilizes the locality constraints projecting each descriptor into its local-coordinate system, which are then integrated by max pooling to generate the final representation, widely used in pattern recognition (Rahmani *et al.* 2014, Yang *et al.* 2016). A fast approximated LLC method is adopted in our experiment, first performing a K-nearest-neighbor search and then solving a constrained least square fitting problem.

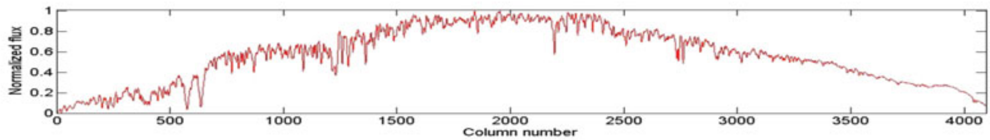
2.4. *Piece-wise optimization*

For a few spectra, the fitting effect of modeling the whole spectrum sometimes is not very good. In consideration of this case, we have tried a piece-wise fitting method, hoping to improve the fitting precision of the spectrum. In our work, first, the spectra are divided

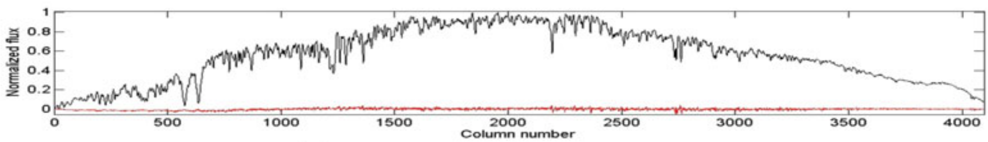


(a) The number of fiber with C > 0.7 (b) The MSE of the modeling results

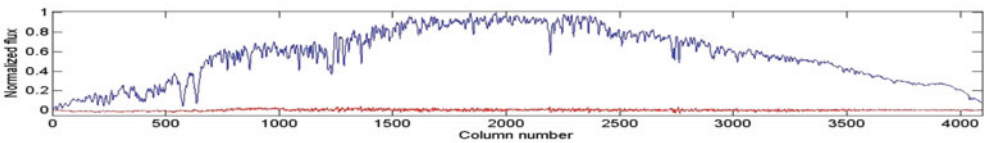
Figure 1. Modeling results based on correlation coefficient



(a) Ground truth



(b) The modeling results by LLC method



(c) The piece-wise modeling results by LLC method

Figure 2. Comparison of the spectrum results

into four parts; then the parameters are optimized by the above method; finally, the four parts are put together to obtain the complete modeled skylight spectra.

3. Results and Conclusion

Fig. 1 (a) shows the number of fiber with correlation coefficient C greater than 0.7, which is expressed as a significant correlation. It can be seen that most of the fibers have a significant correlation; Fig. 1 (b) shows the Mean Square Error (MSE) of the modeling results. The open circles represent the MSE of the distance based method, and the filled diamonds represent the MSE of the correlation coefficient based method. Most of the errors are less than 0.05, and this validates the feasibility of the two modeling methods. And the correlation coefficient based method is much better, because of its smaller error. Fig. 2 (a) shows a skylight spectrum, as a ground truth; Fig. 2 (b) shows the modeling results by LLC method, and the residuals with the ground truth; and Fig. 2 (c) shows the piece-wise modeling results by LLC method and the residuals.

Through these experimental results, we can draw the following conclusions: 1. Utilizing LLC method can well fit the skylight spectra; 2. The error of the spectrum fitting by correlation coefficient based method is smaller than distance based; 3. For the spectra

https://doi.org/10.1017/S1743921316012989 Published online by Cambridge University Press

with smaller correlation coefficient, piece-wise fitting has better effect than the overall fitting.

4. Future

Our current work is preliminary sky background modeling. Therefore, there are still a lot of work to be done:

- In addition to the correlation coefficient and Euclidean distance as the measure of similarity between spectra, we are going to try some other methods;
- Besides the LLC method, there are still a lot of methods for parameter optimization, which is a content that we will further study to find a most suitable method for improving the precision of sky background modeling;
- With the modeled sky background, how to apply it to the object spectral image to achieve two-dimensional sky-subtraction is the key problem that we are going to solve.

Acknowledgment

Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. This research is made use of the data from the National Astronomical Observatories, Chinese Academy of Sciences. The research work described in this paper was fully supported by the Joint Research Fund in Astronomy (U1531242) under cooperative agreement between the NSFC and CAS, the National Natural Science Foundation of China (Project No. 61375045) and Beijing Natural Science Foundation (Project No.4142030). Prof. Fuqing Duan is the author to whom all correspondence should be addressed.

References

Blondin, S., Walsh, J. R., & Leibundgut, B. 2004, *AA*, 431(1), 757
Bolton, A. S. & Schlegel, D. J. 2010, *PASP*, 122(888), 248
Cuby, J. G. 1994, *SPIE*, 98
Cui, X., Su, D., & Wang, Y. 2000, *Proc Spie*, 4003(3), 347
Cui, X. 2006, *Proc Spie*, 6267
Cui, X., Li, G., Zhang, Y., & Li, Y. 2010, *Proc Spie*, 7733(6)
Cui, X., Zhao, Y., Chu, Y., *et al.* 2012, *RAA*, 12(9), 1197
Horne, K. 1986, *PASP*, 98(604), 609
Li, G., Zhang, H., & Bai, Z. 2015, *PASP*, 127(2)
Li, G., Zhang, H., Luo, A., *et al.* 2011, *e-Science Technology & Application*, 2(4)
Piskunov, N. E. & Valenti, J. A. 2002, *AA*, 385(3), 1095
Rahmani, H., Mahmood, A., Huynh, D., *et al.* 2014, *ICPR*, 3511
Robertson, J. G. 1986, *PASP*, 98(609), 1220
Su, D., Cui, X., Wang, Y., & Yao, Z. Q. 1998, *ATI*, 3352, 76
Wang, J., Yang, J., Yu, K., *et al.* 2010, *IEEE*, 3360-3367
Watson, F., Offer, A. R., Lewis, I. J., *et al.* 1998, *FSSR*, 152, 50
Wyse, R. F. G.. & Gilmore, G. 1994, *MNRAS*, 257, 1
Yang, Z. & Xiong, H. 2016, *VC*, 1
Yu, J., Yin, Q., Guo, P., *et al.* 2014, *MNRAS*, 443(2), 1381
Zhao, G., Zhao, Y. H., Chu, Y., *et al.* 2012, *RAA*, 12(7), 723