OUTLIER DETECTION AND RECONSTRUCTION OF LOST LAND SURFACE TEMPERATURE DATA IN REMOTE SENSING

Muhammad Yasir Adnan, Prof Yong Xue and Richard Self

School of Computing & Engineering, University of Derby, United Kingdom

ABSTRACT

In quantitative remote sensing, missing values classified as outliers occur frequently. This is due to technical constraints and the impact of weather on the efficiency of instruments to collect data. In order to deal with these missing values, we offer an Outlier-Search-and-Replace (OSR) algorithm that uses spatial and temporal information for the detection and reconstruction of missing data. The algorithm searches for outlier in the data and reconstruct by finding the best possible match in spatial locations.

KEYWORDS

Remote Sensing, Missing Data Reconstruction, Outlier, MODIS, Land Surface Temperature.

1. INTRODUCTION

The temperature of the terrestrial surface is an essential indicator of the state of the atmosphere. This is extensively used in a wide range of environmental applications, including agriculture. Conditions in the atmosphere have a significant impact on the ability of remote sensing sensors to gather information. For gathering information on the atmosphere, ocean, and land surface, these instruments are the most often used way of data collection. Outlier Search and Replace (OSR) is a technique for detecting and reconstructing outliers in land surface temperature data that is presented in this article. Land surface temperature data from the Moderate Resolution Imaging Spectroradiometer (MODIS) collected in January 2018 is being used for the experiments. The results show that the suggested approach, which takes advantage of both spatial and temporal information, works well when it comes to detecting and reconstructing missing land surface temperature information.

The remainder of the paper is arranged in the following manner. Section 2 discusses the work that is related to it. Section 3 of this work provides a detailed discussion of the outlier identification and reconstruction method that has been proposed in this study. Section 4 summarises the results of experiments conducted using the OSR model. In the end, Section 5 presents the conclusion.

2. RELATED WORKS

The remote sensing platforms are comprised of the equipment or vehicles that are used to collect data from the field. The earth observation data is highly complicated and susceptible to inaccuracies due to the way it is collected [1]. One of the characteristics of the sensors installed is the time of image accusation, the distance between the object and the sensor, the interval between image accusation and image location, and the range of coverage. The missing values are classified as

David C. Wyld et al. (Eds): SIPP, NLPCL, BIGML, SOEN, AISC, NCWMC, CCSIT - 2022
pp. 197-205, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121317

outliers in this study. There are many different types of missing information, which can be broadly classified into the following categories:

- Sensor Malfunction
- Cloud Obscuration

Anomalies introduce outlier in the form of malicious data which is inconsistent with respect to rest of the data. The Figure 1 demonstrates the presence of outliers.

Sensors play a vital role in remote sensing information gathering. The failure of which leads to missing information. For example 15 of 20 detectors in MODIS Aqua band 6 give malicious readings [2]. Typical example of missing data recovery caused by sensor failure is presented by [3]. Sensor malfunction leads to the phenomenon of striping in remote sensing images. An example of striping in a remote sensing images is shown in Figure 2 [4].

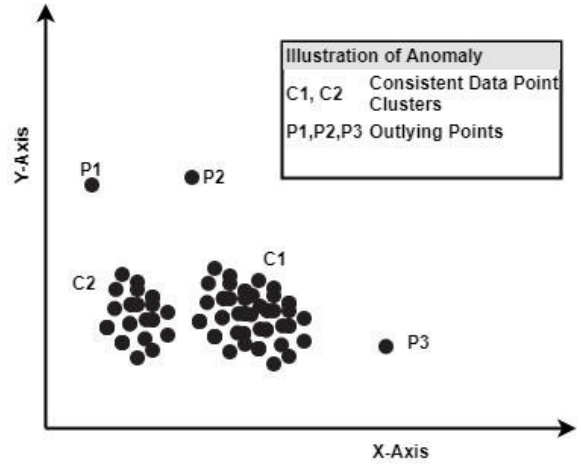

Figure 1. Illustration of Anomaly

Figure 2. Stripping in Remote sensing imagery

Cloud covers hinders the information captured in remote sensing images which greatly reduces the data usability in subsequent application. At one time, 35% of earth surface is covered by clouds [5] and if an individual country is considered, e.g Canada has 50% to 80%of its land is covered with cloud in the morning [3]. Landsat ETM+ scenes are 35% contaminated by clouds which amounts to significant loss of data [6]

Detection of outlier and reconstruction is mainly classified into four categories [3, 5, 7]

- Spatial Based Methods
- Spectral Based Methods

- Temporal Based Methods
- Hybrid Methods

Image inpainting is the essence of spatial based methods. These are many traditional methods of image reconstruction used in the area of remote sensing and computer vision. In image inpainting, it usually assumes the fact that the missing information shares similar geographical features and fills the information gaps with this idea [8]. No Auxiliary images are required for spatial methods [9]. Spatial methods follows the correlation between local and global information of the image to fill the information gap but as the reference data is not large, spatial methods are mainly suitable for reconstruction of small missing areas and results are not guaranteed for complex terrain.

Spectral bands in remote sensing imagery are correlated and the information from another band can be utilized to reconstruct the missing information which overcomes the lack of prior information problem in spatial based methods. When hyper spectral or multi-spectral images have missing information, they both have bands with complete and missing information so the idea is to utilize the bands with complete information to reconstruct missing information by establishing a correlation between the bands. Spectral methods also known as multi-spectral-complementation methods [10]. For example Aqua MODIS has repeated patters of black strips in its imagery due to sensor malfunction in band 6. The solution to this problem was first proposed by [2] in which auther states that Aqua MODIS band 6 and 7 are correlated with coefficient of 0.9821 and 0.9777. Therefor missing information in band 6 can be recovered from highly correlated band 7.

Thick cloud cover causes all the spectral bands to be contaminated and have missing information in them and also the defective sensors may cause missing information in all the bands for a particular spot. Therefor spectral methods become useless as they are based on spectral correlation which is destroyed after all the bands have missing information. At this point, temporal based methods comes into play. As the clouds are continuously moving and data can be acquired for same region at some other time interval. Determination of time interval is a tricky part in this method as if time interval is large, it will be effected by land cover change, but if land cover is small it will have overlapping clouds in two time slots. Lots of work has been done by researchers in temporal based methods, for example [11]presented a method using local linear histogram matching (LLHM) which required high quality data to function but it ends up giving poor results for heterogeneous landscape [12–14] proposed algorithms to improve radiometric consistency of multi-temporal images for heterogeneous land.

Temporal methods outperform all other methods but due to the limitation of amount of land cover change, temporal methods under-perform. Spatio-temporal fusion based method use data fusion from multiple sources to overcome this limitation. Instead of using one reference image [15] used two reference image in close dates to the target cloudy image. The errors produced by temporal methods due to significant land cover changes are avoided by this idea. This method further uses a residual correction strategy to improve spectral similarity between recovered area and remaining cloud free region. Spatial , temporal and spectral method discussed earlier have

their own strength and weaknesses, STS methods are developed by considering the strengths of all these methods to reconstruct missing information much more efficiently and accurately. [4] came up with a unified model called as spatial-temporal-spectral (STS) model based on deep convolutional neural network. The model not only solve the problem of dead lines in MODIS band 6 but also solves corrector-off problem in Landsat Enhanced Thematic Mapper imaging. It is also able to remove thick clouds and shadows using multi-source data. The method establishes a mapping between missing data and complete data with auxiliary data using a deep CNN. The model uses a residual output to learn the relation between different auxiliary data. These methods are also known as hybrid methods [3]. Missing LST data reconstruction for clear sky conditions can overestimate as compare to the reconstructed data under cloudy conditions [16]. While there is a limitation, there is till enough research gap to produce high quality reconstructed land surface temperature data. The author in [17] proposed a robust gap filling method by fusing MODIS and VIIRS LST data. Most algorithms presented in the literature uses one auxiliary image for reconstruction of data. Our algorithm uses auxiliary information from multiple sources as well as multiple days to detect and reconstruct missing values.

3. MODEL

3.1. Dataset

MODIS LST provides observations for daytime and nighttime. These images are taken from MYD11A1 and MOD11A1 which are MODIS LST products. MOD11A1 is captures data at 10:30AM and PM while MYD11A1 captures data at 01:30AM and PM. The image shown in Figure 3 shows data from January 1st 2018 to January 5th 2018 from both the MODIS products. There is high correla- tion between the data from two MODIS products [18] which indicates that temporal information from similar or multiple satellite product can be used effectively for detection and reconstruction of missing data. The raw data from MYD11 and MOD11A1 is downloaded (https://modis.gsfc.n asa.gov/data/dataprod/) and true color images are obtained by processing in ArcGIS with yellow color depicting the missing values.

3.2. Study Area

The dataset is collected from Moderate Resolution Imaging Spectroradiometer (MODIS) and Beijing–Tianjin–Hebei Figure 4 region is selected for initial experimentation. The target area has been used on numerous occasion for various remote sensing applications. The area is approximately 218000km^2 and is located close to the Northwest Pacific Ocean.

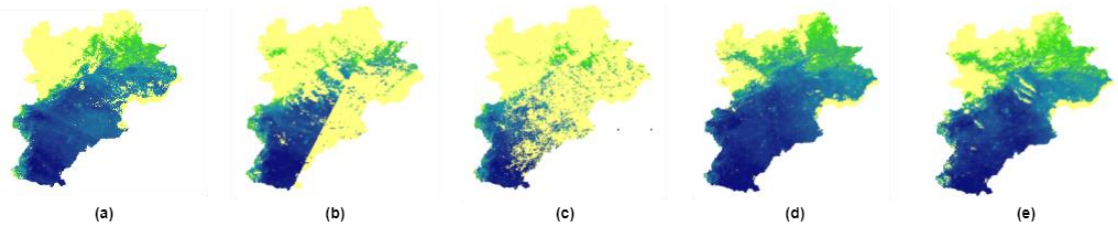

Figure 3. MODIS Input Image Series of 4 Days

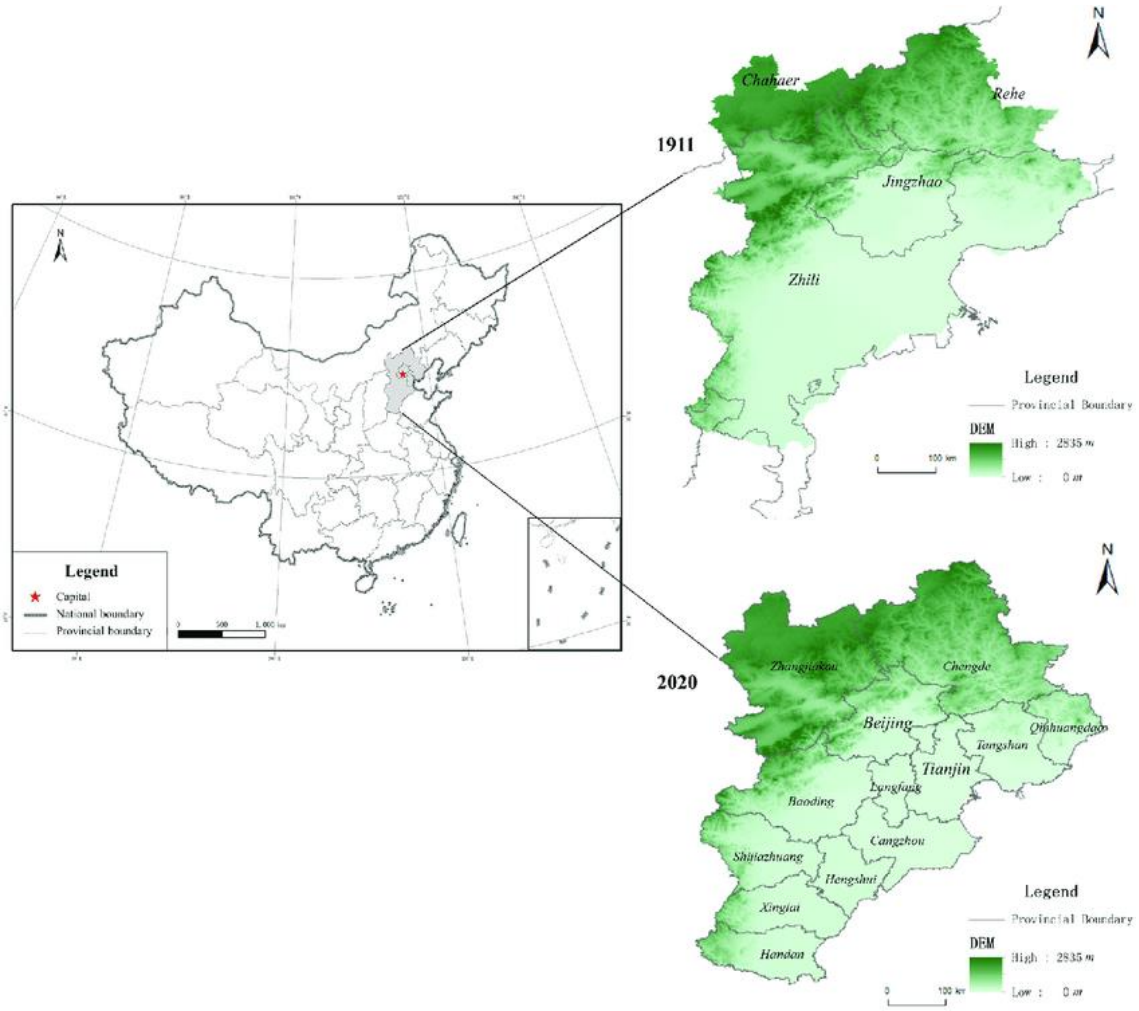

Figure 4. Beijing–Tianjin–Hebei Spatial Location [19]

3.3. Outlier Detection

Region of interest (ROI) of size *MXN* is selected which contains missing values to be detected and reconstructed. Once ROI_w is selected, the same region of interest is selected on all the input images to form a mosaic of temporal information as shown in

Figure 5. Image MOSAIC of M X N windows

Each image slice represent data from one day. Data for each day could be from a different satellite as well with similar temporal and spatial information which makes this algorithm efficient in order to detect and reconstruct missing values.

Dynamic time warping compares each pixel of image mosaic from ROI_w using Equation (1) to form a series of distance values.

$$C_d, D_d = DTW\left(I_{iP(j,k),}\ I_{i+1P(j,k),}\right) \quad (1)$$

$P(j, k)$ is the pixel value at location j, k in each successive day input image region of interest ROI_w. D_d holds distance values between pixels of each successive image slice and $C_d$holds the coordinate of each pixel being compared. As there is high correlation between the temporal LST values of same region in successive days as well as spatial information of data from same satellite as well as multiple satellite [17]. A linear distance curve is formed when pixel values are correct but whenever there is a missing value in the image, the distance value is very high indicating anomaly. These values are identified and located as outliers based on a threshold value T_h and reconstructed in the next section. The threshold value T_h is currently being obtained by experimentation by comparing true pixels and known outliers.

3.4. Missing Data Reconstruction

Once the location of the outlying values C_{d_o} are identified. The reconstruction process begins. The algorithms traverses through the list of distance values D_{d_o} at outlier location C_{d_o} in all temporal image slices in mosaic and finds a pixel value with least distance value using Equation (2).

$$D_d = Min(C_D, D_D)\ \ for\ \ 1 \leq R \leq N \quad (2)$$

C_D in Equation (2) gives the location of the pixel P_{C_D} with lowest distance D_d between pixel of image slices in mosaic at location similar to outlier C_{d_o}. The pixel P_{C_D} is taken as reference pixel to reconstruct the outlier.

Now a $3X3$ window W_o is taken around the location of outlier and the reference pixel P_{C_D} is compared using Equation (2) with the neighboring pixels of the outlying value in W_o. The pixel which is at the lowest distance from the reference pixel is copied at the outlier location. This algorithm is similar to spatial reconstruction of missing values but in this case temporal information is also being utilized to identify the outliers.

4. RESULTS

The images in Figure 3 are used as input to test the accuracy of the algorithm. Outliers were introduced randomly in the input window of $M\ X\ N$ Once the reconstruction process is complete, the reconstructed values are compared with original values before the outliers were introduced. The results are shown in Table 1. The matches show the number of pixels whose values were remained same after the reconstruction and non matches shows number of pixels whose values were different after the reconstruction process which algorithm is finding the best possible match by looking at spatial information in the image.

Table 1. Window Size and number of Random outlying Pixels

Window Size	Number of Random Out-liers	Matches	Non Matches
4 *X* 4	5	1	5
6 *X* 6	10	0	10
8 *X* 8	15	7	8
10 *X* 10	32	24	9
15 *X* 15	62	21	41

The actual image and reconstructed image were compared using Equation (1) and based on matches and non matches and average distance between each of the reconstructed and original pixel value is shows in Figure 6. The average distance was calculated by adding the distance values between actual pixels and reconstructed pixels and dividing by number of respective outliers for window sizes.

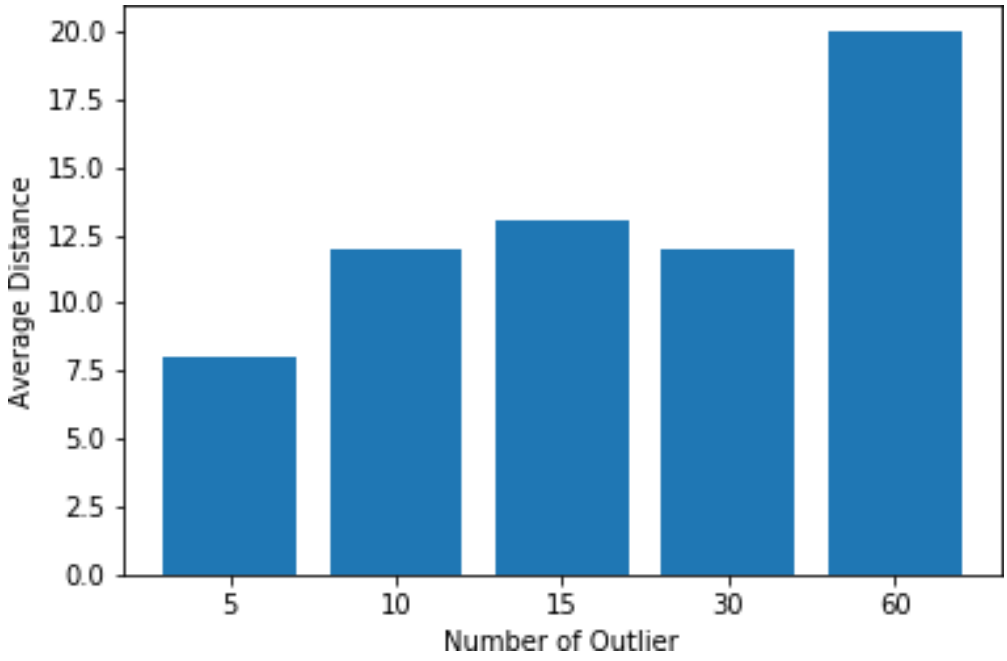

Figure 6. Average Distance between Original and Reconstructed value

The Figure 7 shows pair of image from a-b with input image with outliers and reconstructed image.

5. CONCLUSION

The proposed algorithm for outlier detection and reconstruction in remote sensing makes use of both spatial and temporal information, and as a result, it is referred to as a hybrid algorithm of outlier detection and reconstruction in remote sensing. In order to maximise efficiency in terms of exploiting auxiliary information, the algorithm can make use of temporal information from similar resources as well as from numerous resources at the same time which makes it very flexible.

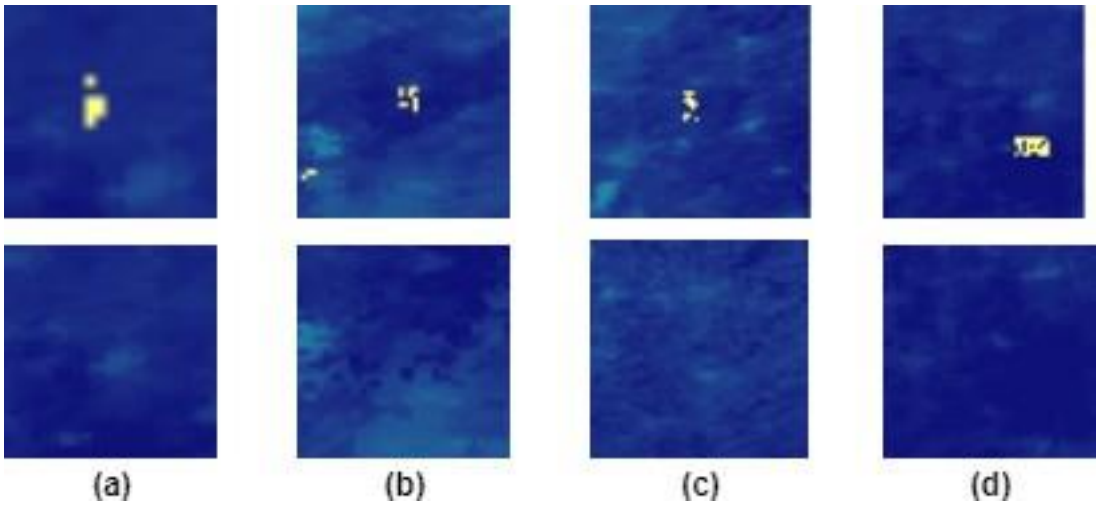

Figure 7. Outlier and Reconstructed Image Pair

REFERENCES

[1] Alexandria Dominique Farias and Gongling Sun, (2020) "Data Mining and Machine Learn- ing in Earth Observation-An Application for Tracking Historical Algal Blooms", in *CS & IT Conference Proceedings*, Vol. 10.

[2] Lingli Wang, John Qu, Xiaoxiong Xiong, Xianjun Hao, Yong Xie, and Nianzeng Che, (05 2006) "A new method for retrieving band 6 of Aqua MODIS", *Geoscience and Remote Sensing Letters, IEEE*, Vol. 3, pp. 267–270.

[3] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, (9 2015) "Missing Information Reconstruction of Remote Sensing Data: A Technical Review", *IEEE Geoscience and Remote Sensing Magazine*, Vol. 3, No. 3, pp. 61–85.

[4] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, (8 2018) "Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 8, pp. 4274–4288.

[5] C. Lin, K. Lai, Z. Chen, and J. Chen, (1 2014) "Patch-Based Information Reconstruction of Cloud-Contaminated Multitemporal Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, No. 1, pp. 163–174.

[6] Junchang Ju and David P. Roy, (2008) "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally", *Remote Sensing of Environment*, Vol. 112, No. 3, pp. 1196–1211.

[7] Wenhui Du, Zhihao Qin, Jinlong Fan, Maofang Gao, Fei Wang, and Bilawal Abbasi, (2019) "An Efficient Approach to Remove Thick Cloud in VNIR Bands of Multi-Temporal Remote Sensing Images", *Remote Sensing*, Vol. 11, No. 11, pp. 1284.

[8] Qiong Lu and Genyuan Zhang, (2018) "Review of Image Inpainting", in *2018 8th Inter- national Conference on Manufacturing Science and Engineering (ICMSE 2018)*. Atlantis Press.

[9] Zhiwei Li, Huanfeng Shen, Qing Cheng, Wei Li, and Liangpei Zhang, (2019) "Thick Cloud Removal in High-Resolution Satellite Images Using Stepwise Radiometric Adjustment and Residual Correction", *Remote Sensing*, Vol. 11, No. 16, pp. 1925.

[10] B. Chen, B. Huang, L. Chen, and B. Xu, (1 2017) "Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 1, pp. 27–37.

[11] James Storey, Pasquale L. Scaramuzza, and Gail Schmidt, (2005) "LANDSAT 7 SCAN LINE CORRECTOR-OFF GAP-FILLED PRODUCT DEVELOPMENT".

[12] Quanjun Jiao, Wenfei Luo, Xue Liu, and Bing Zhang, (2007) "Information reconstruction in the cloud removing area based on multi-temporal CHRIS images", in *MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications*, Yongji

Wang, Bangjun Lei, Jing-Yu Yang, Jun Li, Chao Wang, and Liang-Pei Zhang, Eds. Vol. 6790, pp. 606–612, SPIE.

[13] Bin WANG, Atsuo ONO, Kanako MURAMATSU, and Noboru FUJIWARA, (1999) "Au- tomated Detection and Removal of Clouds and Their Shadows from Landsat TM Images", *IEICE transactions on information and systems*, Vol. 82, No. 2, pp. 453–460.

[14] Din-Chang Tseng, Hsiao-Ting Tseng, and Chun-Liang Chien, (2008) "Automatic cloud removal from multi-temporal SPOT images", *Applied Mathematics and Computation*, Vol. 205, No. 2, pp. 584–600.

[15] Huanfeng Shen, Jingan Wu, Qing Cheng, Mahemujiang Aihemaiti, Chengyue Zhang, and Zhiwei Li, (2019) "A spatiotemporal fusion based cloud removal method for remote sens- ing images with land cover changes", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 12, No. 3, pp. 862–874.

[16] Xiaoma Li, Yuyu Zhou, Ghassem R. Asrar, and Zhengyuan Zhu, (2018) "Creating a seamless 1km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States", *Remote Sensing of Environment*, Vol. 206, pp. 84–97.

[17] Rui Yao, Lunche Wang, Xin Huang, Liang Sun, Ruiqing Chen, Xiaojun Wu, Wei Zhang, and Zigeng Niu, (2021) "A Robust Method for Filling the Gaps in MODIS and VIIRS Land Surface Temperature Data", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15.

[18] (2020) "Filling Gaps of Monthly Terra/MODIS Daytime Land Surface Temperature Using Discrete Cosine Transform Method", *Remote Sensing*, Vol. 12, No. 3.

[19] Shuang Li, Zhongqiu Sun, Yafei Wang, and Yuxia Wang, (08 2021) "Understanding Ur- ban Growth in Beijing-Tianjin-Hebei Region over the Past 100 Years Using Old Maps and Landsat Data", *Remote Sensing*, Vol. 13, pp. 3264.