Framework to Create Cloud-Free Remote Sensing Data Using Passenger Aircraft as the Platform

Wang, Chisheng ; Wang, Shuying; Cui, Hongxing; Šebela, Monja; Zhang, Ce; Gu, Xiaowei; Fang, Xu; Hu, Zhongwen; Tang, Qiandi; Wang, Yongquan

Published in:
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

DOI:
10.1109/JSTARS.2021.3094586

Publication date:
2021

Citation for published version (APA):
Wang, C., Wang, S., Cui, H., Šebela, M., Zhang, C., Gu, X., Fang, X., Hu, Z., Tang, Q., & Wang, Y. (2021). Framework to Create Cloud-Free Remote Sensing Data Using Passenger Aircraft as the Platform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 6923 - 6936. Article 9477015. https://doi.org/10.1109/JSTARS.2021.3094586

Framework to Create Cloud-Free Remote Sensing Data Using Passenger Aircraft as the Platform

Chisheng Wang, Shuying Wang, Hongxing Cui, Monja B. Šebela, Ce Zhang, Xiaowei Gu, Xu Fang, Zhongwen Hu, Qiandi Tang, and Yongquan Wang

***Abstract*—Cloud removal in optical remote sensing imagery is essential for many Earth observation applications. To recover the cloud obscured information, some preconditions must be satisfied. For example, the cloud must be semitransparent or relationships between contaminated and cloud-free pixels must be assumed. Due to the inherent imaging geometry features in satellite remote sensing, it is impossible to observe the ground under the clouds directly; therefore, cloud removal algorithms are always not perfect owing to the loss of ground truth. Recently, the use of passenger aircraft as a platform for remote sensing has been proposed by some researchers and institutes, including Airbus and the Japan Aerospace Exploration Agency. Passenger aircraft have the advantages of short visitation frequency and low cost. Additionally, because passenger aircraft fly at lower altitudes compared to satellites, they can observe the ground under the clouds at an oblique viewing angle. In this study, we examine the possibility of creating cloud-free remote sensing data by stacking multiangle images captured by passenger aircraft. To accomplish this, a processing framework is proposed, which includes four main steps: first, multiangle image acquisition from passenger aircraft, second, cloud detection based on deep learning semantic segmentation models, third, cloud removal by image stacking, and fourth, image quality enhancement via haze removal. This method is intended to remove cloud contamination without the requirements of reference images and predetermination of cloud types. The proposed method was tested in multiple case studies, wherein the resultant cloud- and haze-free orthophotos were visualized and quantitatively analyzed in various land cover type scenes. The results of the case studies demonstrated that the proposed method could generate high quality, cloud-free orthophotos. Therefore, we conclude that this framework has great potential for creating cloud-free remote sensing images when the cloud removal of satellite imagery is difficult or inaccurate.**

***Index Terms*—Cloud removal, deep learning, haze removal, multiple viewing angles, passenger aircraft, photogrammetry.**

Manuscript received April 1, 2021; revised June 2, 2021; accepted June 27, 2021. Date of publication July 7, 2021; date of current version July 22, 2021. This work was supported in part by Chang'an University (Xi'an, China) through the National Key Research and Development Program of China under Grant 2020YFC1512001, in part by the National Natural Science Foundation of China under Grant 41974006, in part by the Shenzhen Scientific Research and Development Funding Program under Grant KQJSCX20180328093453763, Grant JCYJ20180305125101282, and Grant 20200812164904001, in part by the Department of Education of Guangdong Province under Grant 2018KTSCX196, and in part by the Guangdong Special Support Program under Grant 2019BT02H594. *(Corresponding author: Hongxing Cui.)*

Chisheng Wang, Shuying Wang, Xu Fang, Zhongwen Hu, Qiandi Tang, and Yongquan Wang are with the Ministry of Natural Resources (MNR) Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, the School of Architecture and Urban Planning, Shenzhen 518000, China (e-mail: wangchisheng@163.com; wangshuying2019@email.szu.edu.cn; fangxu622@126.com; zwhoo@szu.edu.cn; tangqiandi5105@163.com; wongyq1994@gmail.com).

Hongxing Cui is with the Ministry of Natural Resources (MNR) Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, the School of Architecture and Urban Planning, Shenzhen 518000, China with the South Marine Science and Engineering Guangdong Laboratory, Guangzhou 511458, China, and also with the Department of Ocean Science, Hong Kong University of Science and Technology, Hong Kong (e-mail: hcuiaf@connect.ust.hk).

Monja B. Šebela is with the Sinergise Laboratory for Geographical Information Systems Ltd., 1000 Ljubljana, Slovenia (e-mail: monja.sebela@gmail.com).

Ce Zhang is with the Lancaster Environment Centre, Lancaster University, LA1 4YQ Lancaster, U.K. (e-mail: c.zhang9@lancaster.ac.uk).

Xiaowei Gu is with the Department of Computer Science, Aberystwyth University, SY23 3DB Aberystwyth, U.K. (e-mail: xig4@aber.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2021.3094586

I. Introduction

With the rapid development of remote sensing technology in recent decades, optical remote sensing satellite images have been widely applied in various Earth observation activities, such as climate change assessment, land use and land cover identification, crop mapping, and change detection [1]–[5]. However, cloud coverage is problematic in the retrieval of surface or atmospheric parameters [6], [7], feature extraction [8], and dynamic detection [9] from optical images, the spectral bands of which cover the visible and near-visible wavelengths [10]. With the dramatic increase in remote sensing data obtained from satellites, problematic cloud-contamination in optical remote sensing images has become more apparent. Approximately 67% of the moderate resolution imaging spectroradiometer images are affected by clouds [11]. Cloud coverage blocks the light and obscures the ground surface in remote sensing imagery, therefore, precise identification and removal of cloud coverage are essential for using remote sensing data.

Many cloud removal methods for optical remote sensing data have been presented recently. Traditional methods can be divided into three major categories: Multitemporal [12]–[14], multispectral [15]–[17], and spatial-based approaches [18]–[20]. Multitemporal approaches use temporal images with different acquisition dates to retrieve images without corrupted pixels to yield a cloud-free (CF) image. Lin *et al*. [12] proposed a cloud removal method that uses multitemporal satellite images and information cloning, wherein cloudy areas are cloned from corresponding CF areas based on a global optimization process and the Poisson equation. A nonnegative matrix factorization and error correction method has been used to remove clouds using multitemporal remote sensing data from different sensors

[14]. Additionally, Xu *et al*. [13] introduced a cloud removal method using sparse representation and multitemporal dictionary learning techniques. The above multitemporal approaches are the most prominent techniques in cloud removing. However, CF reference imagery is required for this type of methods, which complicates scene reconstruction due to rapidly changing surface conditions. Multispectral approaches are suitable for the removal of semitransparent clouds and haze. A thin cloud removal approach based on multidirectional dual-tree complex wavelet transform and transfer least square support vector regression was proposed in a previous study [15]. Xu *et al*. [16] developed a thin cloud removal method using signal transmission principles and spectral mixture analysis. These multispectral methods are applied for cloud removal and do not require additional imagery; however, they can only be applied to semitransparent clouds that allow partly spectral transmission of the ground surface. Spatial-based techniques use the hypothetical relationship between contaminated and CF pixels based on spatial and geometric information. A cloud removal method based on similar pixel replacement driven by a spatiotemporal Markov random field model was previously introduced [18]. Meng *et al*. [20] applied a sparse-dictionary-learning-based adaptive patch inpainting method to remove cloud on high-spatial-resolution remote sensing imagery. However, the spatial-based techniques require assuming the spatial relationship between neighboring pixels, which may not stand in many situations. Apart from the three types of traditional methods mentioned above, the deep learning algorithms were recently introduced for cloud removal and show a good performance [21]–[24]. However, deep learning methods require a large number of training samples and the performance may vary significantly in different images.

Although many advanced algorithms are proposed to remove the cloud from satellite remote sensing images, the inherent limitations in satellite platform always make the cloud removal challenging. Satellites are located at significantly high altitudes and the field of view (FOV) is mostly fixed, so the ground under the cloud can hardly be directly observed due to the observing geometry. The inherent limitations are difficult to overcome by any cloud removal algorithms as they all require certain assumptions on the missing information. Recently, using passenger aircraft as the remote sensing platform has been proposed due to the large coverage area, short visitation frequency, and low cost [25]–[27]. Another merit from passenger aircraft is that it can overcome cloud interference due to their flight altitude and multiviewing angles comparing with satellite. Fig. 1 shows that in the image acquisition process, cloud interference increases with satellite platform altitude [28]. However, the passenger aircraft platform has multiview angle observations and can therefore obtain CF surface information, unlike the satellite platform which are fully blocked from ground by the cloud.

Successful earth observation applications using a passenger aircraft platform have been developed [29], [30]. In a previous study [29], a civil aircraft for the regular investigation of the atmosphere based on an instrument container project was used to monitor the atmosphere and was able to provide less costly, real-time meteorological information similar to traditional remote sensing platforms. Passenger aircraft observations have also been adopted for obtaining meteorological data. Recent research finds the COVID-19 pandemic affects the weather forecast as the number of flying passenger aircrafts reduced [30]. Several programs using passenger aircraft as remote sensing platforms have been instituted by different countries globally [31]–[33]. Ray20 uses the Airbus aircraft to build autonomous remote sensing systems over Europe and North America [31]. The Norwegian Research Centre conducted a project that used passenger aircraft equipped with high-resolution imaging systems for environmental monitoring. Additionally, a passenger aircraft used as an observation platform can increase safety and emergency response in the Arctic [32]. These results demonstrate that using passenger aircraft as a remote sensing platform has significant potential for earth observation activities in the future.

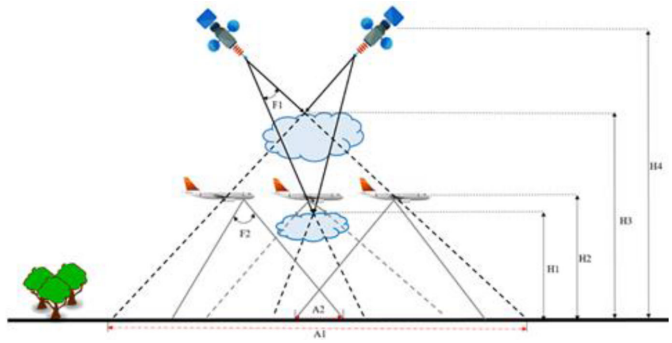


Fig. 1. Map showing cloud impact differences of the satellite and aircraft platforms. A1 is the cloud coverage area using a satellite platform and A2 is the CF area produced using the passenger aircraft platform. H1 is the altitude of the passenger aircraft platform. H2 and H3 are the cloud heights of the different clouds above the earth's surface. H4 is the altitude of the satellite platform. F1 and F2 are the fields of view of the satellite and passenger aircraft, respectively.

Remote sensing has been routinely applied in emergency management such as forest fire detection and flood monitoring [34]. The fast acquisition of remote sensing data is extremely important for initiating effective response [35]. However, the satellite remote sensing imagery cannot provide timely data due to the relatively long revisit period and unpredicted weather condition. On the contrary, the low altitude and large number of commercial flight make it an ideal remote sensing platform for effective emergency response. In this study, a novel framework was developed to generate truly CF orthophotos from a set of time series photos taken with a smartphone camera onboard a passenger aircraft. The proposed method can remove cloud contamination without using information from other images captured in different dime. We implemented the proposed framework in four processing steps. First, a series of images were captured using passenger aircrafts as a platform. Second, deep learning models were adopted to detect clouds. Third, large-scale CF orthophotos were generated through a photogrammetric processing. Finally, the haze-free (HF) orthophotos were obtained using the dark channel prior (DCP) algorithm and histogram statistics. The structure of this study is as follows: Section II describes the proposed method in detail, Section III

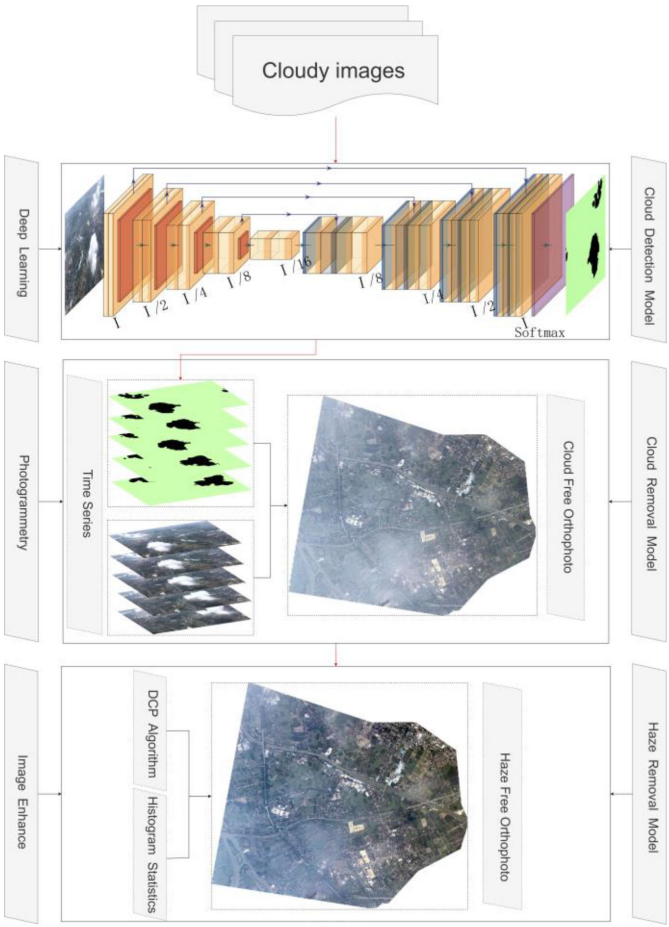


Fig. 2. Flowchart of the orthophoto generation process consisting of cloud detection, cloud removal, and haze removal.

introduces the dataset, Section IV describes the evaluation metrics, and Section V presents the results. Discussions are given in Section VI. Section VII concludes this article.

II. Methods

The framework is composed by data acquisition and data processing. Details about the data acquisition mode using passenger aircraft as the remote sensing platform is presented first in Section II-A. Then the data processing is described, which are performed in three steps: 1) cloud detection using deep learning algorithms, 2) cloud removal from the photos and cloud masks using photogrammetry methods, and 3) haze removal from the orthophotos using image-enhanced methods (Fig. 2). In the first step, three semantic segmentation models are presented, including U-Net (convolutional networks for biomedical image segmentation), feature pyramid network (FPN), and pyramid scene parsing network (PSPNet), which were trained and optimized to assign the appropriate cloud mask to the corresponding images. In the second step, the assigned masks and photos were integrated into the photogrammetry processing to generate a CF orthophoto. In the third step, the haze in the CF orthophoto was removed by using the developed haze removal method that combines DCP algorithm [36] and histogram statistic.

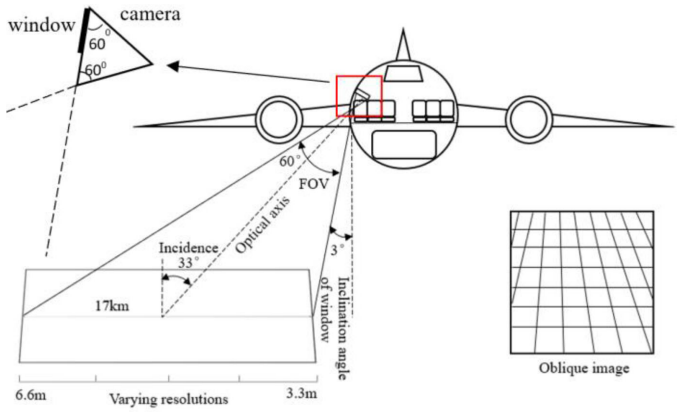


Fig. 3. Imaging geometry using a passenger aircraft as the platform.

A. Passenger Aircraft Data Acquisition

This section briefly describes data acquisition using passenger aircraft as the platform. In typical passenger aircraft, window seats are available from which passengers can capture high-quality pictures of the ground using handheld cameras (e.g., smartphone cameras). In order to obtain overlapping images with high quality, a passenger should take pictures from the window seat near the aircraft tail, as images may be blocked by the wings if passengers sit in the middle of the aircraft. Meanwhile, sky should be avoided in pictures taken. In this study, we sat in the penultimate window seat near the rear of the aircraft. The distance between the lens and the window was kept at about 6 cm. Too short distance may make the lens hard to focus, while too long distance would include the surrounding obstacles into the picture. Meanwhile, the lens was titled to approximately 60° against the window (Fig. 3). If the tilt angle is too small, the sky would appear in raw pictures. Otherwise, the camera would be blocked by the windowsill. During the shooting, the passenger does not need to adjust the camera position. The look angle will change automatically as the aircraft pass the cloud, since the relative position between aircraft position and cloud keeps changing in the flight.

Generally, to meet photogrammetric processing requirements, two essential rules must be implemented when capturing a picture. First, adjacent images with a certain degree of coverage (50%) must be taken to ensure that the pictures can be aligned successfully (Fig. 4). Second, a low incident is required to ensure that more ground details are captured in one picture, as well as to improve dense point matching. Side-look imaging geometry can be obtained when pictures are taken by passengers from both sides of the plane; however, the varying scales of the images are a defect in the oblique images [26], [37]. To obtain high-quality images, the incidence angle should be controlled within a small range. The average FOV of a smartphone camera is approximately 60°. The minimum incidence angle can be adjusted to approximately half the FOV plus the inclination of the seat (approximately 3° measured from a Boeing 787-8 schematic) during a smooth flight. Generally, the altitude of commercial aircraft ranges between 9 and 11 km, the stripe width

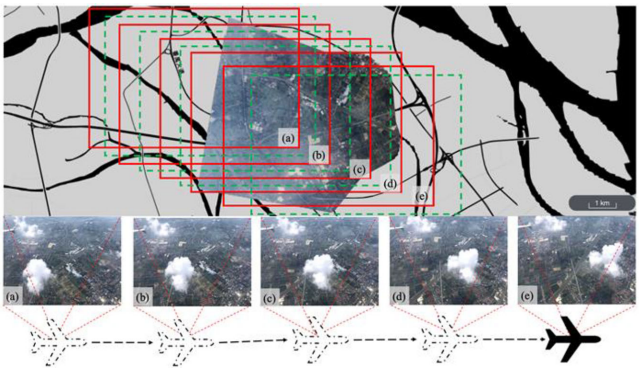


Fig. 4. Example of raw pictures captured by flight CZ3192 during the aircraft's landing. (a) picture ID: IMG_20180805_130702. (b) picture ID: IMG_20180805_1307012. (c) picture ID: IMG_20180805_130722. (d) picture ID: IMG_20180805_130732. (e) picture ID: IMG_20180805_130737. The rectangles with red solid lines all have corresponding raw pictures, the green dashed rectangles without showing raw pictures. The base map includes the cloud-free orthoimage and vector map with geographic information.

is approximately 17 km based on the slant imaging geometry, and the ground resolution ranges between 3.3 and 6.6 m if there are 3000 image pixels along the width (Fig. 3).

B. Cloud Detection Model Architectures

Many deep learning architectures based on semantic segmentation methods have recently been introduced, and some of which are used in satellite imagery [38], [39]. Three representative methods with encoder–decoder structure, i.e., U-Net, FPN, and PSPNet, were trained and tested in the proposed approach. The adopted models were pretrained on the PASCAL VOC-2012 semantic segmentation dataset [40]. To compare the performance of the different methods objectively, three stages of the cloud detection process were investigated. First, the binary cross-entropy function was used to calculate the loss between the predicted and true cloud masks using the different methods. Second, the stochastic gradient descent was selected using momentum [41] as the optimizer in each method's training stage. Third, each method produced a model adapted from the ResNet-101 network [42].

U-Net, which is based on a fully convolutional neural network [43], is built on an encoder–decoder architecture that is comprised of a contracting path to capture context and a symmetric expanding path to enable accurate location. The two main features of the U-Net architecture are the U-shaped network structure and the skip connection. The U-shaped network structure consists of downsampling on the left and symmetric upsampling on the right. To obtain the feature map of the image, the encoder performs feature extraction, which consists of convolution and downsampling, and the fusion method used by the skip connection concatenates on the feature map channels. Finally, the feature map is restored to the original resolution of the corresponding cloud mask based on the downsampling and convolution of the decoder.

The fully convolutional neural network has been further improved and is now known as the FPN [44], which extracts features at different scales to form a pyramidal hierarchy and efficiently uses the semantic information of the different scales in the FPN. The structure of the FPN can be summarized as feature extraction, upsampling, feature fusion, and multiscale feature output. The input and output images of the FPN are feature maps of different scales. The FPN architecture is divided into bottom-up and top-down pathways. The bottom-up pathway is a feature encoder process using the ResNet-101 network. The top-down pathway with upsampling and lateral connections builds high-level semantic feature maps at different scales using the corresponding cloud masks.

The PSPNet architecture also contains an encoder and a decoder. Specifically, the encoder contains a pyramid pooling module and a convolutional neural network [45] backbone with dilated convolutions instead of the fully convolutional neural network. The dilated convolution layers can capture a more receptive field and the pyramid-pooling module is used for capturing the global context from an input image, which helps the PSPNet network to classify the pixels from the global information present in the image. After the encoder features of the image are extracted, the decoder takes the features and converts them into feature representations using upsampling and concatenation layers. Finally, the representation is fed into a convolutional layer to obtain the corresponding per-pixel cloud mask.

C. Cloud Removal Model

Considering the characteristics of the obtained images, such as the oblique acquisition angle and uneven illumination of the images caused by the different acquisition angles, a traditional photogrammetric method is not suitable for generating the orthophoto. However, the varied viewing angle geometry at which the cloudy and CF images of the same area are obtained is important because CF orthophoto generation is possible by combining these multiangled photos (Fig. 5). To overcome obstacles such as the oblique acquisition angle and uneven illumination of the images, a processing procedure for CF orthophoto generation is proposed in this study, which consists of the following four steps.

1) Camera Position Initialization: Approximate camera positions are necessary for rapid conjugate point detection and georeferencing of the sparse point cloud. Additionally, the cloud masks detected in the images are not calculated during conjugate point detection. Because aircraft GPS information is obtained by commercial companies and publicized for flight tracking purposes, the flight GPS positions can be obtained from the flight tracking information. As passenger aircraft are relatively stable, we can use piecewise linear interpolation to obtain continuous GPS positions of a flight with an interval in seconds (Fig. 6). Specifically, the variable y here refers to altitude, longitude, or latitude. Therefore, a linear interpolation of these three variables can be done, respectively, along the time dimension to obtain the corresponding GPS parameters at each second. The continuous GPS positions information as a reference, further, coordinate information in pictures taken were extracted at the corresponding time. It is worth noting that by setting the interpolation value to the initial value of the camera position, the positioning error of

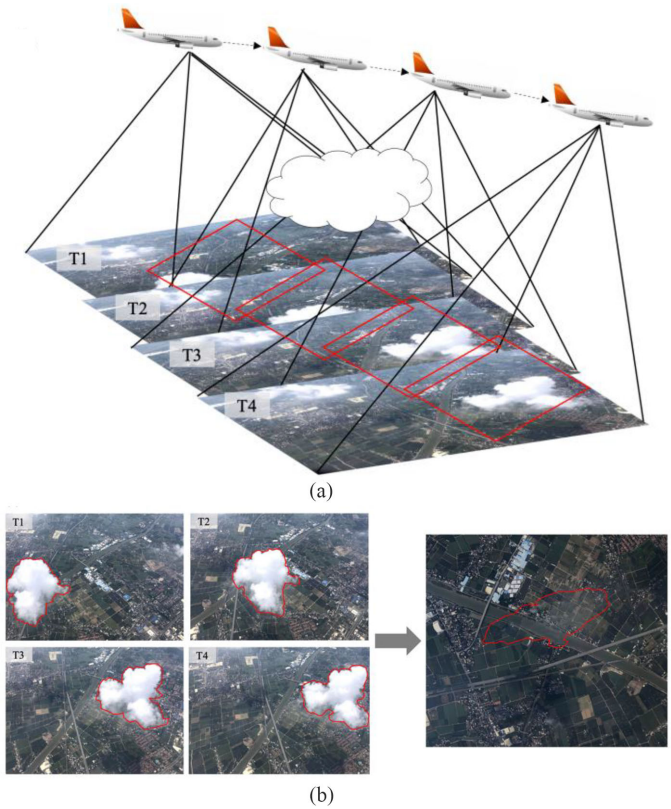


Fig. 5. Cloud removal performed using images produced by multiple viewing angles (a) diagram illustrating how time-series images are obtained and (b) multiview oblique images and the orthophoto after cloud removal.

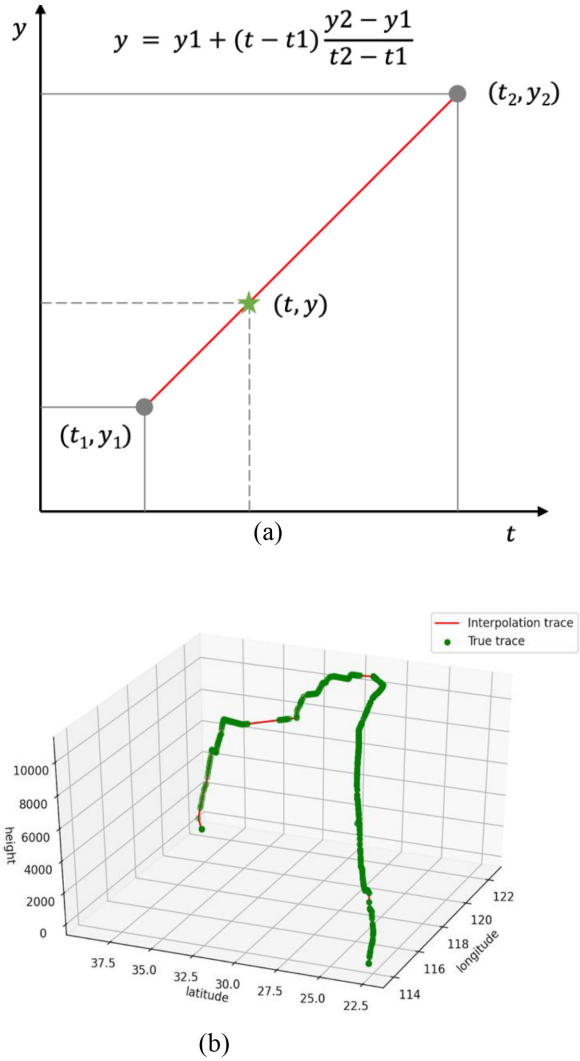


Fig. 6. (a) Schematic of 1D piecewise linear interpolation along time. (b) Piecewise linear interpolation of the discrete aircraft positions downloaded from FlightAware. Green points are the original flight tracking points with position information, and red line is the flight CZ3192 trace interpolated from original flight tracking points.

the ground control points (GCPs) would be reduced in a further step. Our previous research has shown that the measurement accuracy of the GPS was roughly set to be about ten times greater than that of the ground control points (GCPs), with a general ground sampling resolution of less than 6 m [26].

2) Interior Orientation Parameters Initialized from the Exchangeable Image File Header: In addition to the camera positioning information, internal positioning parameters such as focal length and sensor size are required to facilitate conjugate point searching. The interior camera orientation parameters are further refined in Step 3.

3) Structure From Motion (SfM) Processing: SfM is a 3D reconstruction algorithm based on the initialized GPS information and interior orientation parameters [46]. SfM processing can be performed by many aerial photogrammetry software programs (e.g., Agisoft Metashape, Pix4d, Menci APS, and MicMac). On comparing the performances of these software in processing a large number of images, the results showed that Metashape provides acceptable accuracy and satisfactory computational performance with graphics processing unit acceleration.

4) Orthophoto Mosaic Generation: A dense point cloud can be generated when the SfM stage is completed. Meanwhile, a digital surface model can be obtained based on the regular interpolation of the point cloud. The input photos are orthorectified to orthophotos using the individual camera positions and digital surface model. To generate mosaic CF orthophotos, several strategies can be used, including averaging the values of all pixels from the individual photos, taking pixels from the photo observations closest to the normal direction, and using frequency domain approach. The coordinate system for the mosaic orthophoto can be set arbitrarily as required.

D. Haze Removal Model

Haze is a common characteristic in remote optical sensing images. Commercial aircraft typically fly between 9 and 11 km above the surface, therefore, objects observed at this height will be obstructed by different cloud types such as cumulus, altocumulus, and stratus, and haze generally occurs with certain cloud types. If haze is treated as a cloud to generate a larger number of cloud masks, a large portion of the resultant orthophoto will be missing. In this study, to obtain a complete orthophoto, a haze mask was not used as a cloud mask in the cloud detection process. Instead, a haze removal method was developed that

TABLE I
FLIGHT INFORMATION

Flights No.	Flight route	Date	Time	Images in dataset 1	Images in dataset 2
CZ3192	BJ-SZ	20180805	10:08-12:59	10	10
HU7703	BJ-SZ	20190808	08:10-11:35	10	12
CZ6591	SZ-NB	20190623	09:00-10:29	12	17

combined the DCP algorithm and histogram statistics based on the characteristics of the orthophoto. The primary reason for employing this method is to use the DCP algorithm to remove haze from the orthophoto and then use the histogram statistics to restore the color of the HF orthophoto.

The DCP algorithm [34] indicated that at least one color channel has very low intensity or approaches 0 for some pixels in most of the nonsky patches of the image. This algorithm can be divided into three steps. First, the dark channel is calculated and the atmospheric light intensity is estimated. Subsequently, a refined transmission function with soft matting is obtained. Finally, the atmospheric light intensity is calculated. The HF image can be reconstructed based on the unknown values calculated. To restore the color of the orthophoto, the histogram method was used to enhance the HF orthophoto quality.

III. EXPERIMENTAL DATA

We collected a set of cloudy images from three flights (Table I). Their image capturing methods are similar. Therefore, we took flight 1 as an example for data acquisition description. It is a flight from China Southern Airlines (CZ3192) which took off from Beijing International Airport (Code: PKX) at 10:23 AM local time and lands at Shenzhen Ba'oan International Airport (Code: SZX) at 13:20 PM on August 5, 2018. The aircraft used for the flight is the Airbus A330, which is a medium-size, wide-body aircraft, capable to continuously fly 13 450 km with 247 passengers. The aircraft has 58.82-m length and 17.39-m height. Fig. 7 shows the flight trace information of the flight and the location of pictures taken by the camera on that trace. We started to collect data at altitude of about 4700 m. Flight information above was obtained from FlightAware website (https://zh.flightaware.com/). Pictures were acquired with iPhone X camera from 13:07 PM to 13:08 PM local time. Taking adjacent pictures with too long interval would produce inconsistent orthoimage. To ensure that cloud and aircraft have not moved too much to be far apart, each picture were taken at five-second intervals, and a total of ten cloudy pictures with a high overlapping rate were acquired (Fig. 3). Each raw picture is composed of three channels of red, green, and blue, with a size of 4032 × 3024, and has time information about moment of capture.

The experimental data were divided into two parts. The first part was used for training and validation based on deep learning methods for clouds detection. Based on the passenger aircraft platforms, we have collected a set of cloudy images during two years (from 2018 to 2019) by taking into account weather

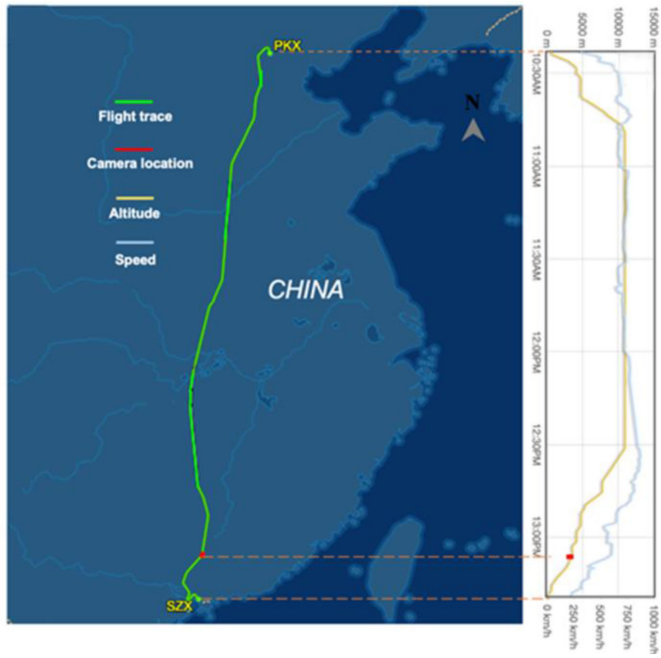


Fig. 7. Flight information for flight CZ3192 (from FlightAware).

conditions, type of clouds, as well as characteristics in ground objects, depending on available images, We randomly divided 32 images into two data sets, 20 images for training and 12 images for validation. The cloud mask labeling procedure was performed using 32 images following three steps. First, the cloudy image was stretched into a proper visual contrast using Adobe Photoshop. Subsequently, the magic wand tool was used to mark the cloudy locations in the image. Finally, a manually labeled reference mask was generated by assigning the cloud and CF pixel values of 0 and 255, respectively [47]. To expedite training process, both training images and corresponding cloud masks were divided into multiple nonoverlapping 500 × 500 patches, and then 3788 image patches and 3788 mask patches were input into a cloud detection model network. The testing dataset was also processed in the same way to extract 2552 patches from images and masks, respectively.

The second part of the dataset was used to predict cloud masks and generate the corresponding orthophotos. To evaluate the quality of orthophotos generated under different environmental conditions, the prediction dataset from three representative scenes was selected based on cloud type, haze density, and vegetation coverage, and images for two of the scenes were captured during flight CZ6591 on June 23, 2019. During this flight, a set of 17 time-series images of Shenzhen City (herein referred to as scene1) and a set of 12 images of Huizhou City (herein referred to as scene2) were taken at a 5-s time interval between adjacent images. During flight CZ3192 on August 5, 2018, a set of 10 time-series images were taken of Guangzhou City (herein referred to as scene3). Evaluating the quality of the CF orthophotos based on the various cloud types present at the time they were taken facilitated an understanding of the effect of cloud characteristics on CF orthophotos. Haze is frequently found in remote sensing images, therefore, haze density

was considered during this evaluation, which contributed to obtaining comprehensive cloud removal knowledge in a specific scene. The spectral feature is one of the most important features of cloud detection [48]. The normalized green–red difference index (NGRDI), with values ranging between -1 and 1, was selected as the spectral feature [49] in this study because the image channels only had RGB bands. Then the vegetation cover density can be calculated by combining dimidiate [50] pixel model and NGRDI, the vegetation cover density and NGRDI are defined as follows:

$$\mathrm{NGRDI} = [\mathrm{Green\ DN} - \mathrm{Red\ DN}] \,/\, [\mathrm{Green\ DN} + \mathrm{Red\ DN}] \tag{1}$$

where the DN represents a digital number in each pixel, Green DN and Red DN represent DN in green channel and red channel, respectively. If the denominator is 0, the corresponding value of NGRDI is set as 0

$$F = [\mathrm{NGRDI} - \mathrm{NGRDIsoil}] \,/\, [\mathrm{NGRDIveg} -\!- \mathrm{NGRDIsoil}] \tag{2}$$

where the *F* represents the vegetation cover density, NGRDIsoil and NGRDIveg represent the vegetation index when the vegetation density is 0% and 100%, respectively. If the denominator is 0, the corresponding value of *F* is set as 0.

According to the literature [50], the representative scenes were classified into four categories: bare soil areas (BSA), low-level vegetation cover (LVC), medium-level vegetation cover (MVC), and high-level vegetation cover (HVC) with F thresholds of (0, 10%), (10%, 25%), (25%, 50%), and (50%, 100%), respectively. Scene1, scene2, and scene3 had F values of 13.8%, 27.6%, and 56.7% and were therefore classified as LVC, HVC, and MVC, respectively.

IV. Evaluation Metrics

To evaluate the performance of the cloud detection results and image quality after haze removal objectively, different quantitative indicators were selected.

Because the cloud masks obtained via U-Net, FPN, and PSP-Net used different datasets, the predicted masks were compared against the corresponding ground truth masks and classified as "cloud" (positive) or "clear" (negative). The datasets were evaluated quantitatively based on the metrics of accuracy, recall, precision, F-score, and Jaccard index. Accuracy is defined as the percentage of accurately predicted masks in the total sample, which can be used as an indicator for evaluating the accuracy of different models. However, accuracy is not an objective indicator to evaluate the performance of models when the data types are unbalanced. Precision is defined as the number of classified clouds that were literally clouds. Recall is defined as the number of cloud pixels that were classified. The F-score provides insight into the optimum balance between recall and precision, and the Jaccard index is a measure of the similarity between the truth masks and predicted masks [51]–[53]. These five metrics are calculated as follows:

$$\text{Overall Accuracy} = \frac{\sum_{i=1}^{M}(tp_i + tn_i)}{\sum_{i=1}^{M}(tp_i + tn_i + fp_i + fn_i)} \tag{3}$$

$$\text{Precision} = \frac{\sum_{i=1}^{M} tp_i}{\sum_{i=1}^{M}(tp_i + fp_i)} \tag{4}$$

$$\text{Recall} = \frac{\sum_{i=1}^{M} tp_i}{\sum_{i=1}^{M}(tp_i + fn_i)} \tag{5}$$

$$\text{F1score} = \frac{2}{\text{Recall}^{-1} + \text{Precision}^{-1}} \tag{6}$$

$$\text{Jaccard Index} = \frac{\sum_{i=1}^{M} tp_i}{\sum_{i=1}^{M}(tp_i + fp_i + fn_i)} \tag{7}$$

where *tp, tn, fp*, and *fn* are the numbers of true positive, true negative, false positive, and false negative pixels in each test image, respectively, and M denotes the total number of images in each test dataset.

In this study, two full-reference metrics [54] and no-reference image quality assessment (IQA) models were used to fully evaluate the HF orthophotos. The full-reference metrics used were the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The PSNR is calculated by the error of the corresponding pixels, and a large PSNR value indicates small distortion [55], [56]. The SSIM is used to measure image similarity based on brightness, contrast, and structure [57], and a large SSIM value indicates less image distortion. The no-reference image evaluation method was the blind/referenceless image spatial quality evaluator (BRISQUE) in which the mean subtracted contrast normalized coefficients, and neighborhood coefficients are fitted with the generalized and asymmetric generalized Gaussian distribution models, and then the image quality is evaluated using these model parameters [58]. The two full-reference IQA metrics are defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{255}{\sqrt{|x_{in} - x_{out}|^2}} \tag{8}$$

where x_{in} and x_{out} represent the HF and dehazed images, respectively

$$\mathrm{SSIM}x_{in},\ x_{\mathrm{out}} = \frac{(2\mu_{x_{in}}\mu_{x_{out}} + \theta_1)(2\sigma_{x_{in}x_{out}} + \theta_2)}{(\mu_{x_{in}}^2 + \mu_{x_{out}}^2 + \theta_1)(\sigma_{x_{in}}^2 + \sigma_{x_{out}}^2 + \theta_2)} \tag{9}$$

where $\mu_{x_{in}}$ and $\mu_{x_{out}}$are the averages of x_{in} and x_{out}, respectively, $\sigma_{x_{in}}^2$ is the variance of x_{in}. $\sigma_{x_{\mathrm{out}}}^2$is the variance of x_{out}. $\sigma_{x_{in}x_{\mathrm{out}}}$ is the covariance with x_{in} and x_{out};, θ_1 and θ_2 are constants used to avoid system instability caused by a denominator of 0.

V. Results

A. Cloud Detection Results

The quantitative accuracy evaluation results of cloud detection are presented in Table II, and the qualitative evaluation of the cloud masks generated based on various cloud detection methods is presented in Fig. 8.

Among the three ground vegetation cover types, the evaluation indices of the three cloud detection models were highest with MVC. According to the results in Fig. 8(a), the strong

TABLE II
EVALUATION RESULTS FOR THE CLOUD DETECTION METHODS OF PERCENT ACCURACY OF PREDICTED DATA FOR THE THREE SCENES

Dataset (no. scenes)	Method	Overall acc.	Recall	Precision	F-score	Jaccard
MVC (10 images)	U-Net	98.87	97.48	**96.71**	97.09	94.44
	FPN	98.84	97.57	96.49	97.02	94.32
	PSPNet	**98.89**	**97.74**	96.58	**97.15**	**94.56**
LVC (17 images)	U-Net	97.80	94.08	**95.25**	94.65	90.14
	FPN	**97.81**	**94.19**	95.17	**94.67**	**90.17**
	PSPNet	97.67	93.81	94.89	94.34	89.61
HVC (12 images)	U-Net	94.66	91.71	92.29	92.00	85.51
	FPN	94.97	**91.96**	92.92	92.43	86.24
	PSPNet	**95.17**	91.51	**93.87**	**92.62**	**86.56**

Bold values denote the highest accuracy for each scene among the different cloud detection methods.

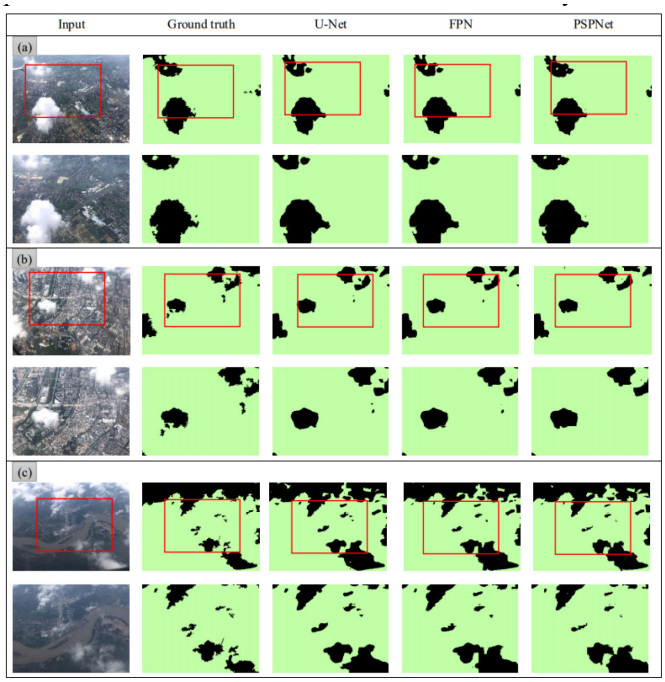


Fig. 8. Examples of cloud and cloud mask detection results using the three cloud detection methods on the three different scenes. (a) MVC. (b) LVC. (c) HVC.

contrast between the clouds and the surrounding objects shows that in certain areas, thick regular shaped clouds completely occluded ground objects, making it easy for the model to mask the clouds. Additionally, the haze densities in the sky were low. These environmental factors provided good conditions for the cloud detection model architectures to learn local and global features from the MVC image. Table I shows that all the cloud detection models performed well in each evaluation index, with PSPNet achieving the highest accuracy, recall, F-score, and Jaccard values and U-Net achieving the highest precision value (96.71%). For the same evaluation index, the difference between the results of the methods was less than 0.3%, which means all the methods had high cloud detection accuracy in MVC.

Fig. 8(b) shows that the LVC scene is primarily covered by urban areas, has minimal vegetation, and the radiation intensities of some of the buildings and clouds are similar, therefore, the contrast between them is poor. Additionally, the cloud boundaries are blurred, and the thin cloud cover is fragmented. These environmental factors create difficulties for cloud detection models to differentiate between clouds and ground objects in the extraction of global and local features. The haze density in the sky, however, is the lowest among the three scenes, which contributes positively to training the model architecture. In the LVC scene, FPN achieved the highest values for each evaluation index, except precision, which was only 95.17%. The highest precision value of 92.25% was achieved by U-Net. Notably, the difference in the calculation results of the various methods was less than 0.5% when comparing the same evaluation index, indicating that the three cloud detection models showed good consistency with regard to the various accuracy evaluation results.

Most areas in the HVC scene [Fig. 8(c)] were covered by natural vegetation, except for the river basin, and the high-density vegetation contributed to a strong contrast between ground objects and clouds, which is conducive to accurate cloud detection. However, haze densities were high, which reduced the contrast between the clouds and other features, and there were a large number of fragmented clouds, which caused difficulties in determining the thin cloud boundaries. Additionally, several pixels containing both ground objects and extremely thin clouds were semitransparent. These environmental factors are extremely unfavorable for thin cloud detection. PSPNet achieved the highest accuracy, recall, F-score, and Jaccard values, and FPN achieved the highest precision of 91.96%. Based on these results, the cloud detection accuracy was the lowest for the HVC case, and the differences in evaluation index values between the three methods were greater than 1%, which shows that the performance of the three methods was not stable compared to the performance of the LVC and MVC scenes.

B. CF Orthophoto Results

To obtain a better understanding of the spatial visualization of the HF orthophotos using the different cloud detection methods, orthophotos of the MVC, LVC, and HVC scenes were used for subjective visual evaluation (Fig. 9).

In the MVC scene, most of the cloud boundaries were distinct, and the contrast between ground objects and clouds was evident [Fig. 8(a)]. These factors are favorable for the extraction of cloud features using deep-learning-based cloud detection methods, therefore, highly accurate cloud mask results were generated by the three cloud detection methods, and the evaluation results were similar (Table I). The quantitative results were consistent with what can be observed, although a few thin clouds remained in the orthophotos [yellow circles in Fig. 9(a)] because they were difficult to separate from the surrounding objects. Moreover, some pixels containing both clouds and ground objects were semitransparent. The generated orthophoto would contain fewer thin clouds if the semitransparent pixels had been classified as nonclouds for the training and testing procedures of the cloud detection methods. If the semitransparent pixels had been classified as clouds, the orthophoto would return no values in these areas.

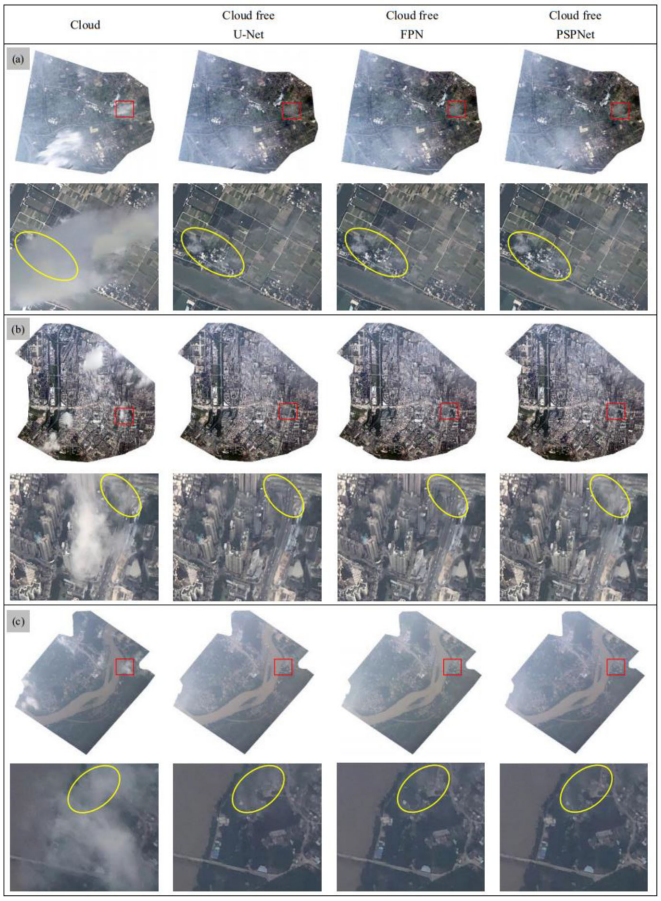


Fig. 9. Qualitative comparison of CF orthophotos generated by different cloud detection methods for various scenes (a) MVC, (b) LVC, and (c) HVC. The cloudy and corresponding CF locations are indicated by yellow circles.

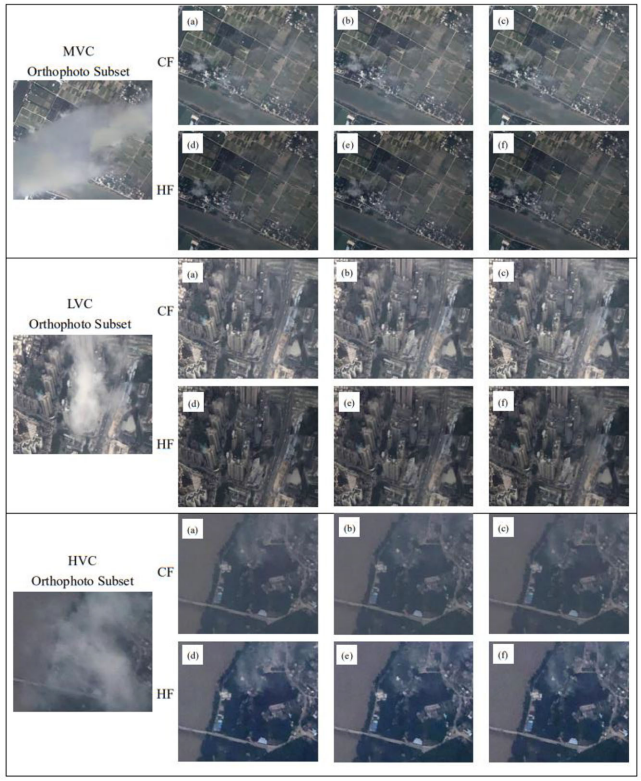


Fig. 10. Qualitative comparison of the CF and HF orthophoto results obtained by the cloud and haze removal methods, respectively, of the MVC, LVC, and HVC scenes using the different cloud detection methods. The cloud removal results using (a) U-Net, (b) FPN, and (c) PSPNet and the haze removal results using (d) U-Net, (e) FPN, and (f) PSPNet.

To obtain a comprehensive orthophoto, some semitransparent pixels had to be classified as nonclouds in this study.

For the LVC scene, the high reflectivity of buildings reduced the contrast between the clouds and ground objects and blurred some thin cloud boundaries [Fig. 8(b)]. The cloud detection accuracy was lower than the MVC scene, and the results of PSPNet were not as accurate as other cloud detection methods (Table I). The differences between the methods are evident from the visual differences in the orthophoto subset [yellow circles in Fig. 9(b)]. A few blurry thin clouds found in the subset are associated with PSPNet, which classified many semitransparent pixels as nonclouds [Fig. 8(b)]. Therefore, the U-Net and FPN cloud detection results, which were similar, were better than the PSPNet results because they classified the same subset area as no-cloud. This indicates the high accuracy of the cloud detection results, which was consistent with the visual orthophotos without clouds in the LVC scene.

The HVC scene had the highest haze densities compared with other scenes, and the cloud fragmentation was the highest and had many semitransparent pixels [Fig. 8(c)]. These specific factors are detrimental to accurately capturing cloud features, hence, the performances of the models were the lowest among the scenes (Table I). The high incidence of semitransparent pixels increased the uncertainty when matching feature points between two adjacent images, and due to the high haze density, the visual performance of the HVC orthophoto was worse than those of the MVC and LVC orthophotos (Fig. 9). There were significant differences in the quantitative evaluation metrics produced by the three cloud detection methods and PSPNet performed better than U-Net and FPN. Additionally, the visual performance of the orthophoto subset [yellow circles in Fig. 9(c)] could not be easily distinguished by the cloud detection methods.

C. HF Orthophoto Quality Assessments

In this section, to evaluate the image quality of HF orthophotos objectively and comprehensively, HF and CF orthophoto subsets (obtained via the haze and cloud removal methods, respectively) of the MVC, LVC, and HVC scenes, using the different cloud detection methods, were compared. The results are presented in Fig. 10, and the corresponding quantitative results comparing the three described IQA metrics are shown in Table II.

The visual examples of the three scenes show that the details of the HF orthophoto subset are more refined and display ground objects more clearly. No significant differences in the HF results could be observed between the different cloud detection methods. The results of the CF orthophoto subset are consistent with those of the HF results. By comparing the different scenes,

better image quality enhancement occurred in the HVC scene compared with the other two scenes.

The visual results are consistent with the statistical results, where all the IQA metrics values are similar for the same scene. For the PSNR and SSIM metrics, the MVC orthophoto subset yielded the best results, which indicates that the CF and HF data are the closest and have the lowest haze densities. The PSNR and SSIM results were better in the LVC scene than the HVC scene, indicating that the haze densities in the LVC scene were lower than those in the HVC scene, which concurs with the visual results between these scenes (Fig. 10). Obvious differences in the PSNR and SSIM metrics existed between the PSPNet results and the U-Net and FPN results. This is attributed to fewer thin clouds in the orthophoto subset generated by PSPNet in the LVC scene. According to the no-reference IQA results, the MVC data yielded the best BRISQUE results that were similar to the corresponding HVC results, while the statistical results of the LVC were the worst. Because the BRISQUE was used to generate the HF orthophoto without reference to the CF orthophoto, the shapes of the ground objects were simple and regular with few color distortions in the MCV areas and, the BRISQUE had excellent performance. In the LVC scene, the buildings made feature extraction difficult, the generated orthophoto was slightly distorted, and a large number of shadows were present in the scene. Consequently, the BRISQUE result was worse for the LVC scene than those for the other scenes.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT SCENES AND MODELS BY IQA METRICS

Scenes	Methods	PSNR	SSIM	BRISQUE
MVC	U-Net	25.569	0.951	30.012
	FPN	25.388	0.939	30.023
	PSPNet	**25.621**	**0.962**	**30.437**
LVC	U-Net	18.122	0.826	35.481
	FPN	**18.297**	**0.889**	**35.053**
	PSPNet	15.146	0.688	37.142
HVC	U-Net	15.195	0.859	31.782
	FPN	15.241	0.865	31.182
	PSPNet	**15.460**	**0.890**	**31.039**

Bold values denote the highest IQA metrics

VI. DISCUSSION

We found that cloud detection performance of deep learning-based methods is significantly influenced by dataset size, and environmental conditions (e.g., cloud characteristics, haze densities, and types of ground objects). Three deep learning models were used to detect clouds with the same dataset. It was challenging for us to obtain large contaminated datasets with clouds under various scenes in a short time. Therefore, the small dataset in this study may limit the capacity of deep learning models to learn global and local features in the image. Besides, we found a cloud detection model has different detection accuracy in different scenes, which may be directly related to environmental conditions. Previous studies have also presented that the same model may perform well for thick clouds [37], [51] but relatively worse for a scene with snow cover on the ground [59]. However, these environmental factors interference for cloud detection tasks are inevitable. With the development of deep learning technology and the increase of dataset size, these issues would also be solved to a certain extent.

The quality of the generated CF orthophoto was affected by two factors, including the cloud detection accuracy and the environmental conditions (e.g., haze densities). First, a satisfactory cloud detection accuracy is critical for cloud removal. Specifically, if there are some cloud pixels misclassified into noncloud during the cloud detection process, they may act as feature points in the photogrammetry processing and appear in the final CF orthophoto. It is the reason why some thin clouds remain in the orthophoto (Fig. 8). However, if the noncloud pixels were incorrectly classified as cloud pixels, the orthophoto may have data missing over there due to the lack of ground objects information. Second, the performance of cloud removal also depends on the environmental conditions. In scene with intensive haze densities, the accuracy of feature point matching in overlapping images would be reduced. However, the feature matching is a key step in the photogrammetric processing for generating orthophotos [60]. The inaccurate feature matching would induce slight distortion in ground objects and destroy the image quality. Therefore, an accurate cloud detection and appropriate environmental conditions is necessary for generating high-quality CF orthophotos.

In the haze removal, the combined method of DCP and histogram statistics was used to enhance the orthophoto qualities. The details in the HF orthophoto were significantly improved comparing with the CF orthophoto. The quality of the final HF orthophoto is also related to the cloud detection accuracy and environmental conditions. Evaluated by the IQA metrics, we found obvious differences existed in the final HF orthophotos with different cloud detection model and different scenes. By comparing the cloud detection accuracy Table II and the image quality evaluation statistical result Table III, it can be seen that the cloud mask detection accuracy is consistent with IQA metrics results for the same scene. We concluded that the image quality in the HF orthophoto has been significantly improved comparing with the CF orthophoto.

Furthermore, to clarify the advantages of cloud removal using passenger aircraft as platforms in cloudy weather, we created a visualized map based on a Landsat 8 scene and obtained two CF orthophotos from flight CZ6591, wherein the Landsat 8 scene overlaid the orthophotos [Fig. 11(a)]. Clouds in the Landsat 8 scene completely covered the LVC and surrounding areas [Fig. 11(b)]. Under these conditions, the contaminated area cannot be reconstructed using spatial-based and multispectral approaches. Spatial-based methods require a hypothetical relationship between cloudy and CF pixels [18], [19], and multispectral methods require semitransparent cloud or haze conditions [15], [17]. Landsat 8 images only covered half of

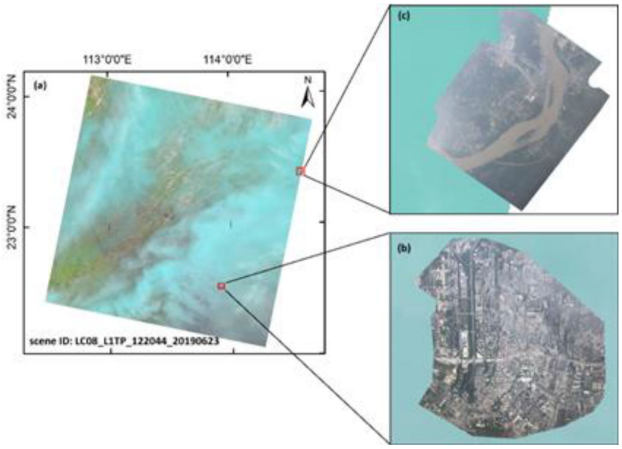


Fig. 11. Two orthophotos from flight CZ6591 and the overlaid Landsat 8 scene taken on the same day. Zoom images of the (a) LVC and (b) HVC areas.

the HVC areas. Such conditions limit the use of spatial-based and multispectral methods, as well as multitemporal approaches with a reference CF image [13]. This is because additional CF images are taken from the same satellite at the same position and recovered after at least 1 revisit period. It is difficult to meet the requirements necessary to perform quality dynamic monitoring of the Earth's surface activities. Deep-learning-based cloud removal approaches [22], [23] have large uncertainties during the learning process if a satellite image is completely contaminated by clouds, therefore, only using cloud removal algorithms on fully contaminated satellite images obtained by optical sensors on remote sensing platforms cannot generate truly CF images.

However, there are also some inevitable limitations and challenges in the proposed method. Compared to professional digital aerial cameras, consumer-grade cameras have smaller sensor sizes and produce images of relatively lower quality. It is difficult to obtain the corresponding orientation and positioning parameters from pictures taken based on aircraft platforms. As far as the oblique observation direction is concerned, there are some defects in the acquired images, such as inconsistent scale, the presence of obstacles, and the existence of invisible areas. Due to the complexity of cloud types and environmental factors, there were still some errors in the detection of clouds. The factors above would have different degrees of influence on the quality of the generated CF orthophotos. The potential applications may be affected to some degree due to these limitations: 1) For quantitative remote sensing applications, the final orthoimages may not meet the requirement for precise parameter inversion, 2) the inconsistent quality require more attentions to be paid during the data processing. In future work, we can integrate high-quality satellite image to refine the geometry and radiation quality, and therefore partly overcome these limitations.

In addition, the size of observed area from an individual flight is much smaller than satellite. Fortunately, more than 100 000 flights per day worldwide provide the possibility of observing large-scale area based on passenger aircrafts. In order to estimate the area that can be covered by global flights per day, trends in the number of global flights were analyzed based on the flight tracking statistics data from Flightradar24

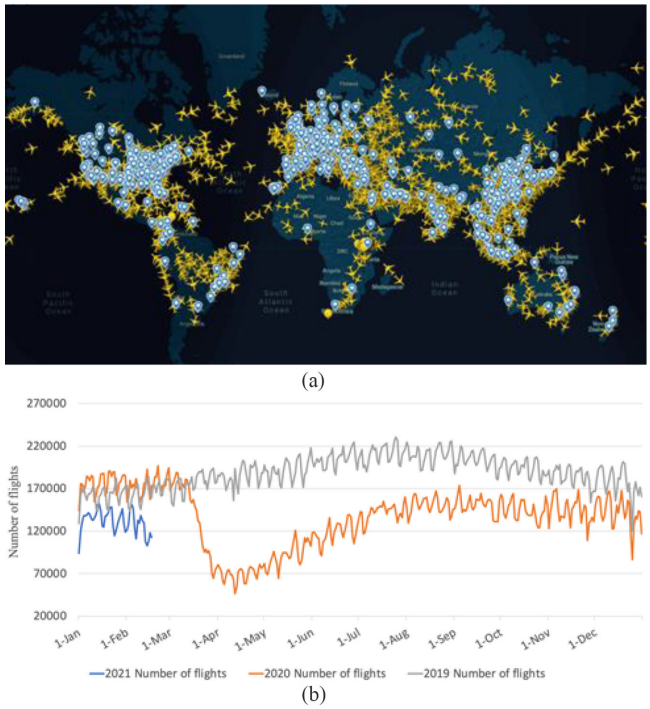


Fig. 12. (a) Global flying aircrafts distribution at 7:20 on 2021-02-17. Blue pins denote the airport locations. (b) Total number of flights tracked by Flightradar24, per day (UTC time), 2019 vs 2020 vs 2021.

(https://www.flightradar24.com/). Civil aviation routes connect about 200 countries and regions and 1700 airports [Fig. 12(a)]. The overall trend for the average number of flights operating per day worldwide in 2019 was relatively stable, all above 100 000 daily flights [Fig. 12(b)]. We can find a significant downward trend in the number of daily flights worldwide starting in April 2020 due to the global spread of the COVID-19. Nevertheless, the average number of daily flights for the whole year was still about 10 000. According to the study by Wang et al. (2020), assuming an average speed of 900 km/h and a strip width of 34 km, with 8000 aircraft flying over our heads every moment, the area covered per day is 5.8750×10^9 km^2, which is close to the size of the earth's surface [26]. In fact, there are many areas that cannot be covered by flights, which can be complemented by satellite imagery, and to a certain extent, the two observation methods can complement each other very well.

With the coverage of more flight routes around the world, a larger-scale earth observation based on passenger aircraft platforms can be even more realistic. This work can be further refined in future by considering the time-series images under various scenes, including cloud coverage at different altitudes, clouds mixed with snow, cloud shadows, as well as Sun patterns, e.g., day/night. After collecting more data, we can integrate state-of-the-art deep learning-based cloud detection models with advanced photogrammetry and computer vision techniques to create high-quality HF orthophotos in multiple scenes.

VII. Conclusion

Optical satellites are inevitably hindered by clouds when obtaining remote sensing images. All cloud removal methods require certain assumptions on blocked pixels, as there is hardly direct observation over there from satellite. The cloud coverage in satellite imagery is associated with the altitude and FOV of the satellite platform. The extent of cloud contamination in satellite imagery is more evident as the satellite's altitude increases. Additionally, the optical sensor on the satellite platform captures the surface information at the same angle due to its fixed FOV, causing information located in the same place to be always contaminated by clouds in sequential remote sensing images. Therefore, generating a truly CF image from a fully contaminated image is difficult using only cloud removal approaches.

Compared with the FOV and altitude limitations of the satellite platform, the passenger aircraft platform has the advantages of suitable altitude and multiviewing angles to capture ground-visible images in cloudy conditions. In this study, a framework was presented to generate a CF orthophoto using passenger aircraft as the remote sensing platform. The proposed method can combine images from multiple viewing angles and remove cloud contamination without distinguishing cloud types or the need for a reference image. This study presented the cloud removal results from three representative scenes using the proposed framework, which considered cloud type, haze density, and vegetation coverage. The clouds in the orthophotos of the different scenes were all effectively removed. The image quality of the CF orthophotos was highly associated with the accuracy of detected cloud masks, which were significantly influenced by cloud characteristics, haze density, and image contrast. However, we found there were no obvious differences among the three cloud detection methods (U-Net, FPN, and PSPNet) in detecting cloud. By including the haze removal step, the quality of the CF orthophotos is significantly improved. Our study demonstrates that the framework can create high quality, and truly CF orthophotos. It can be used to generate rapid and uncontaminated remote sensing product in some particular applications like emergency response and disaster monitoring.

References

[1] W. Kalisa *et al.*, "Assessment of climate impact on vegetation dynamics over East Africa from 1982 to 2015," *Sci. Rep.*, vol. 9, 2019, Art. no. 16865.

[2] A. Fisher, N. Flood, and T. Danaher, "Comparing Landsat water index methods for automated water classification in eastern Australia," *Remote Sens. Environ.*, vol. 175, pp. 167–182, 2016.

[3] C. Zhang, P. A. Harrison, X. Pan, H. Li, I. Sargent, and P. M. Atkinson, "Scale sequence joint deep learning (SS-JDL) for land use and land cover classification," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111593.

[4] A. Asokan and J. Anitha, "Change detection techniques for remote sensing applications: A survey," *Earth Sci. Inform.*, vol. 12, pp. 143–160, 2019.

[5] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111425.

[6] J. Wei *et al.*, "Satellite-derived 1-km-resolution PM1 concentrations from 2014 to 2018 across China," *Environ. Sci. Technol.*, vol. 53, pp. 13265–13274, 2019.

[7] J. Wei *et al.*, "Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach," *Remote Sens. Environ.*, vol. 231 2019, Art. no. 111221.

[8] T. Bai, D. Li, K. Sun, Y. Chen, and L. Wenzhuo, "Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion," *Remote Sens.*, vol. 8, pp. 715, 2016.

[9] S. Mahajan and B. Fataniya, "Cloud detection methodologies: Variants and development—A review," *Complex Intell. Syst.*, vol. 6, pp. 251–261, 2020.

[10] J. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, pp. 1196–1211, 2008

[11] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, Jul. 2013.

[12] C. Lin, P. Tsai, K. Lai, and J. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.

[13] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, "Cloud removal based on sparse representation via multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2998–3006, May 2016.

[14] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 148, pp. 103–113, 2019.

[15] G. Hu, X. Li, and D. Liang, "Thin cloud removal from remote sensing images using multidirectional dual tree complex wavelet transform and transfer least square support vector regression," *J. Appl. Remote Sens.*, vol. 9, no. 1, 2015, Art. no. 095053.

[16] M. Xu, M. Pickering, A. J. Plaza, and X. Jia, "Thin cloud removal based on signal transmission principles and spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1659–1669, Mar. 2016.

[17] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS J. Photogrammetry Remote Sens.*, vol. 149, pp. 215–225, 2019.

[18] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 92, pp. 54–68, 2014.

[19] L. Lorenzi, F. Melgani, and G. Mercier, "Missing-Area reconstruction in multispectral images under a compressive sensing perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3998–4008, Jul. 2013.

[20] F. Meng, X. Yang, C. Zhou, and Z. Li, "A sparse dictionary learning-based adaptive patch inpainting method for thick clouds removal from high-spatial," *Resolution Remote Sens. Imagery*, vol. 17. 2017, Art. no. 2130.

[21] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.

[22] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 1726–1729.

[23] J. D. Bermudez, P. N. Happ, R. Q. Feitosa, and D. A. B. Oliveira, "Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1220–1224, Aug. 2019.

[24] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 333–346, 2020

[25] C. Wang, J. Ke, W. Xiu, K. Ye, and Q. Li, "Emergency response using volunteered passenger aircraft remote sensing data: A case study on flood damage mapping," *Sensors (Basel)*, vol. 19, 2019, Art. no. 4163.

[26] C. Wang *et al.*, "Volunteered remote sensing data generation with air passengers as sensors," *Int. J. Digit. Earth*, vol. 14, no. 2, pp. 158–180, 2021.

[27] T. Mastelic, J. Lorincz, I. Ivandic, and M. Boban, "Aerial imagery based commercial flights as remote sensing platform," *Sensors (Basel)*, vol. 20, no. 6, 2020, Art. no. 1658.

[28] S. Dehnavi *et al.*, "Cloud detection based high resolution stereo pairs geostationary meteosat images," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 371.

[29] R. J. Fleming, "The use of commercial aircraft as platforms for environmental measurements," *Bull. Amer. Meteorological Soc.*, vol. 77, no. 10, pp. 2229–2242, 1996.

[30] Y. Chen, "COVID-19 pandemic imperils weather forecast," *Geophysical Res. Lett.*, vol. 47, no. 15, 2020, Art. no. e2020GL088613.

[31] "Maturing computer vision and connectivity solutions within swarm networks for autonomous earth observation and GIS," Acubed, Sunnyvale, CA, USA, 2020. [Online]. Available: https://acubed.airbus.com/projects/ray-20/
[32] "Norway: World's first passenger aircraft for environmental monitoring," Norwegian Res. Centre, Bergen, Norway, 2019. [Online]. Available: https://www.norceresearch.no/en/insight/verdens-forste-passasjerfly-for-miljoovervaking-er-norsk
[33] "AXA and A. N. A. launch Joint, research on remote sensing observation of atmospheric components using passenger aircraft and satellites—-Contributing global warming measures from sky and space." Jpn. Aerosp. Exploration Agency, Tokyo, Japan, 2020. [Online]. Available: https://global.jaxa.jp/press/2020/09/20200928-1_e.html
[34] S. Voigt, T. Kemper, T. Riedlinger, R. Kiefl, K. Scholte, and H. Mehl, "Satellite image analysis for disaster and crisis-management support," *IEEE Trans. Geosci. remote Sens.*, vol. 45, no. 6, pp. 1520–1528, Jun. 2007.
[35] Y. Ma, F. Chen, J. Liu, Y. He, J. Duan, and X. Li, "An automatic procedure for early disaster change mapping based on optical remote sensing," *Remote Sens.*, vol. 8, no. 4, pp. 272, 2016.
[36] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[37] S. Verykokou and C. Ioannidis, "Oblique aerial images: A review focusing on georeferencing procedures," *Int. J. Remote Sens.*, vol. 39, pp. 3452–3496, 2018.
[38] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, pp. 247–259, 2019
[39] C.-C. Liu *et al.*, "Clouds classification from Sentinel-2 imagery with deep residual learning and semantic image segmentation," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 119.
[40] M. Everingham *et al.*, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vision*, vol. 111, pp. 98–136, 2015.
[41] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1600.04747*.
[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2016, pp. 770–778.
[43] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[44] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc.#IEEE Conf. Comput. Vision Pattern Recognit.*, Jul. 2017, pp. 936–944.
[45] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
[46] M. J. Westoby, J. Brasington, N. F. Glasser, M. J. Hambrey, and J. M. Reynolds, "Structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications," *Geomorphology*, vol. 179, pp. 300–314, 2012.
[47] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS J. Photogrammetry Remote Sens.*,, vol. 150, pp. 197–212, 2019.
[48] Q. Xiong *et al.*, "A cloud detection approach based hybrid multispectral features with dynamic thresholds for GF-1," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 450.
[49] E. R. Hunt, M. Cavigelli, C. S. T. Daughtry, J. E. McMurtrey, and C. L. Walthall, "Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status," *Precis. Agriculture*, vol. 6, pp. 359–378, 2005.
[50] M. Zribi, S. Le Hégarat-Mascle, O. Taconet, V. Ciarletti, D. Vidal-Madjar, and M. R. Boussema, "Derivation of wild vegetation cover density in semi-arid regions: ERS2/SAR evaluation," *Int. J. Remote Sens.*, vol. 24, pp. 1335–1352, 2003.
[51] S. Mohajerani, T. A. Krammer, and P. Saeedi, "A cloud detection algorithm for remote sensing images using fully convolutional neural networks," in *Proc. IEEE 20th Int. Workshop Multimedia Signal Process.*, Aug. 2018, pp. 1–5, doi: 10.1109/MMSP.2018.8547095.
[52] S. Mohajerani and P. Saeedi, "CPNet: A context preserver convolutional neural network for detecting shadows in single RGB images," in *Proc. IEEE 20th Int. Workshop Multimedia Signal Process.*, Aug. 2018, pp. 1–5, doi: 10.1109/MMSP.2018.8547080.
[53] S. Mohajerani and P. Saeedi, "Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Jul./Aug. 2019, pp. 1029–1032.
[54] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of quality measures," *J Electron Imag.*, vol. 11, pp. 206–223, 2002.
[55] D. Jobson, Z.-U. Rahman, G. Woodell, and G. Hines, "A comparison of visual statistics for the image enhancement of FORESITE aerial images with those of major image classes," *Proc. SPIE - Int. Soc. for Opt. Eng.*, 2006, Art. no. 624601.
[56] S. Shao, Y. Guo, Z. Zhang, and H. Yuan, "Single remote sensing multispectral image dehazing based on a learning framework," *Math. Problems Eng.*, vol. 2019, 2019, Art. no. 4131378.
[57] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[58] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[59] S. Qiu, Z. Zhu, and B. He, "Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111205
[60] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7thIEEE Int. Conf. Comput. Vision.*, Sep. 1999 vol. 1152, pp. 1150–1157.

Chisheng Wang received the B.S. degree in GIS from Beijing Normal University, Beijing, China, in 2007, the M.S. degree in cartography and GIS from the Institute of Applied Remote Sensing, Chinese Academy of Science, Beijing, in 2010, and the Ph.D. degree in geodesy from the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

He is currently an Associate Professor with the School of Architecture & Urban Planning, Shenzhen University, Shenzhen, Guangdong, China. His research interests include image processing and remote sensing applications.

Shuying Wang received the B.S. degree in geographic information science from Jiangxi Normal University, China, in 2019. She is currently working toward the M.S. degree in geographic information and smart cities from Shenzhen University, Shenzhen, China.

Her research interests include GIS, InSAR, and remote sensing.

Hongxing Cui received the B.S. degree in food science and engineering from Guangxi University of Science and Technology, Liuzhou, China, the M.Sc. degree in environmental science and engineering from Shanghai Ocean University, Shanghai, China. He is currently working toward the Ph.D. degree in marine environmental science from Hong Kong University of Science and Technology, Hong Kong.

His research interests include marine environment, climate change, remote sensing, and machine learning.

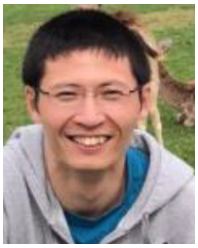
Monja Blatnik Šebela received the B.S degree in philosophy from the University of Ljubljana, Ljubljana, Slovenia, where she is currently working toward the M.Sc. degree in geography.

She is currently working as a GIS analyst with Sinergise Laboratory for Geographical Information Systems, Ljubljana, Slovenia, focusing on satellite processing and visualization, creating scripts displaying various phenomena on the ground. Her research interests include remote sensing, GIS, data fusion, pedology, physical geography, design, and coding.

Ce Zhang received the Ph.D. degree in geography from Lancaster Environment Centre, Lancaster University Lancaster, U.K., in 2018.

He is currently a Lecturer of Geospatial Data Science with the Centre of Excellence in Environmental Data Science (CEEDS), joint venture between Lancaster University and U.K. Centre for Ecology & Hydrology (UKCEH). His research interests include geospatial artificial intelligence, machine learning, deep learning, and remotely sensed image analysis.

Dr. Zhang was the recipient of the prestigious European Union (EU) Erasmus Mundus Scholarship for a European Joint M.Sc. program between the University of Twente, Enschede, The Netherlands, and the University of Southampton, Southampton, U.K.



Xiaowei Gu received the Ph.D. degree in computer science from Lancaster University U.K., in 2018 and the M.Eng. degree in communication and information systems and the B.Eng. degree in communication engineering from Hangzhou Dianzi University, China, in 2012 and 2008.

He is currently a Lecturer in Computer Science at the Department of Computer Science, Aberystwyth University, U.K. His major research interests include machine learning, data analytics, and signal processing.

Zhongwen Hu received the B.Sc. degree in remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

He is currently an Associate Professor with Shenzhen University, Shenzhen, China. His research interests include object-based image analysis and coastal remote sensing.

Qiandi Tang received the B.S. degree in communications and transportation from Shenyang Jianzhu University, Shenyang, China, and the M.Sc. degree in traffic and transportation engineering from Shenzhen University, Shenzhen, China.

She is currently a Presales Engineer with Guangzhou Dushijuan Network Technology Co., Ltd, Guangzhou, China. Her research interests include image processing and night-time light remote sensing applications.

Yongquan Wang received the B.S. degree in engineering surveying from Anhui Agricultural University, Hefei, Anhui, China, in 2017 and the M.S. degree in urban-informatics in 2021 from Shenzhen University, Shenzhen, Guangdong, China, where he is currently working toward the doctoral degree with the School of Architecture & Urban Planning in Remote Sensing.

His research interests include image processing and remote sensing applications.