Article

Forest Types Classification Based on Multi-Source Data Fusion

Ming Lu [1,2] (ID), **Bin Chen** [3] (ID), **Xiaohan Liao** [1], **Tianxiang Yue** [1,2,*], **Huanyin Yue** [1],
Shengming Ren [4], **Xiaowen Li** [3], **Zhen Nie** [5] and **Bing Xu** [3,5,*]

[1] State Key Laboratory of Resources and Environment Information System,
Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences,
Beijing 100101, China; lum@lreis.ac.cn (M.L.); liaoxh@igsnrr.ac.cn (X.L.); yuehy@lreis.ac.cn (H.Y.)
[2] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Department of Earth System Science, Tsinghua University, Beijing 100084, China;
chenbin@mail.bnu.edu.cn (B.C.); xiaowenli627@126.com (X.L.)
[4] Key Laboratory of Watershed ecology and Geographical Environment Monitoring,
National Administration of Surveying, Mapping and Geoinformation, Nanchang 330209, China;
renshengming@mrl.org.cn
[5] State Key Lab of Remote Sensing Science, College of Global Change and Earth System Science,
Beijing Normal University, Beijing 100875 China; niezhen9876@mail.bnu.edu.cn
* Correspondence: yue@lreis.a.cn (T.Y.); bingxu@tsinghua.edu.cn (B.X.);
Tel.: +86-010-64889041 (T.Y.); +86-010-62790189 (B.X.)

Abstract: Forest plays an important role in global carbon, hydrological and atmospheric cycles and provides a wide range of valuable ecosystem services. Timely and accurate forest-type mapping is an essential topic for forest resource inventory supporting forest management, conservation biology and ecological restoration. Despite efforts and progress having been made in forest cover mapping using multi-source remotely sensed data, fine spatial, temporal and spectral resolution modeling for forest type distinction is still limited. In this paper, we proposed a novel spatial-temporal-spectral fusion framework through spatial-spectral fusion and spatial-temporal fusion. Addressing the shortcomings of the commonly-used spatial-spectral fusion model, we proposed a novel spatial-spectral fusion model called the Segmented Difference Value method (SEGDV) to generate fine spatial-spectra-resolution images by blending the China environment 1A series satellite (HJ-1A) multispectral image (Charge Coupled Device (CCD)) and Hyperspectral Imager (HSI). A Hierarchical Spatiotemporal Adaptive Fusion Model (HSTAFM) was used to conduct spatial-temporal fusion to generate the fine spatial-temporal-resolution image by blending the HJ-1A CCD and Moderate Resolution Imaging Spectroradiometer (MODIS) data. The spatial-spectral-temporal information was utilized simultaneously to distinguish various forest types. Experimental results of the classification comparison conducted in the Gan River source nature reserves showed that the proposed method could enhance spatial, temporal and spectral information effectively, and the fused dataset yielded the highest classification accuracy of 83.6% compared with the classification results derived from single Landsat-8 (69.95%), single spatial-spectral fusion (70.95%) and single spatial-temporal fusion (78.94%) images, thereby indicating that the proposed method could be valid and applicable in forest type classification.

Keywords: data fusion; forest types; classification

1. Introduction

Forests are among the most biologically-diverse and largest terrestrial ecosystems on Earth [1]. They play an important role in global carbon and hydrological cycles and provide a wide range

of valuable ecosystem goods and services, such as food, timber and climate moderation [2,3]. High-accuracy forest mappings including the types, spatial distribution, canopy structure, tree species composition and temporal changes are of great importance to forest management, conservation biology and ecological restoration. Forests can be classified in different ways and to different degrees of specificity. Forest types are defined as a group of forest ecosystems with a generally similar composition that can be differentiated from other groups by their species composition, productivity or crown closure [4]. The distinction is whether the forests are composed predominantly of broad-leaved trees, coniferous trees or mixed. Identification of forest types at fine resolution is critical to provide useful information for forest managers, as well as ecological modelers [5].

Forest inventories are regarded as the most frequent way to obtain forest properties' information with the highest accuracy. However, the traditional field survey approach is time consuming and labor intensive. Participatory sensing/citizen science has become a new cost-effective way to collect in situ forest data [6,7], but participatory sensing approaches cannot solve all the problems, especially in isolated areas where few people reach. The use of remote sensing data is still the best way to obtain accurate and timely forest information over large spatial scales and long-term temporal coverages. Currently, remote sensing of forests has developed into a new discipline. On the one hand, it improves our understanding of how and why remotely-sensed data and methods are important in forestry and forest science. On the other hand, it in turn strengthens our awareness that a better understanding of forest ecosystems may be essential for harmonized coexistence between humans and nature [8].

A number of previous studies has presented different methods to differentiate forest types using various remote sensing data. Ren et al. [9] used a hierarchical classification method to distinguish different forest types in complex mountainous areas by incorporating high spatial resolution remote sensing images and multi-source auxiliary data. Torresan et al. [10] exploited metrics extracted from an airborne LIDAR (Light Detection and Ranging) raw point cloud to predict different forest structure types by means of classification trees. Gorgens et al. [11] utilized Airborne Laser Scanning (ALS) to discriminate different Brazilian forest types based on canopy height profiles, which revealed that it was possible to differentiate forest types using canopy height profiles derived from ALS data. Chen et al. [12] proposed a spatial feature extraction method that used the Vegetation Local Difference Index (VLDI) derived from the Normalized Difference Vegetation Index (NDVI) to increase the accuracy of forest type classification. The results showed that combining the spatial information extracted from medium-resolution images and spectral information improved both classification accuracy and visual quality. Castilla et al. [13] harmonized four independent land cover datasets and different satellite images (SPOT, Landsat and MODIS) to produce a common and simple forest map consisting of three classes of forest (needle-leaf, broad-leaved and mixed) and non-forest. Connette et al. [14] used multi-spectral Landsat OLI imagery for delineating main forest types in Myanmar's Tanintharyi Region and estimated the extent of degraded forest for each unique forest type. These studies have covered most of the commonly-used remote sensing data sources including active LIDAR, SAR, multispectral, hyperspectral, thermal systems, etc. In addition, they also have covered most of the widely-used classification methods, such as K-means, ISODATA, maximum likelihood, the spectral angle mapper, Bayesian, Support Vector Machine (SVM), neural network, random forest, etc. Despite many advances having been achieved in these researches to perform accurate forest type classification, most of the existing fine-resolution forest-type classification methods are determined by the availability of very high-resolution images and the incorporation of complex physical models associated with specific forest types. However, these datasets and methods seem to present difficulty for widespread use.

Although there has been a growing number of satellites launched over the past few decades, the trade-off among spectral resolutions, spatial coverages and repeat frequencies still cannot be properly solved. So far, no single satellite sensor can generate images of fine spatial, temporal and spectral resolutions [15]. However, the spatial, temporal and spectral resolutions represent the ability of presenting details of the Earth's surface, repeated observation and spectral detection, respectively, which are all vital indicators to identify different forest types. Fortunately, multi-source

data fusion breaks through the constraints of a single sensor and effectively integrates the advantages of multiplatform complementary observations, thus providing opportunities to achieve more accurate and comprehensive forest classification and monitoring [16]. Many data fusion methods have been proposed over the past few decades. Shen [17] proposed the integrated fusion method to obtain the complementary information from multiple temporal-spatial-spectral images; however, the temporal, spatial and spectral characteristics of objects were not completely considered. Zhang et al. [18] proposed a method called Ratio Image-Based Spectral Resampling (RIBSR), which is used to accomplish data resampling in the spectral domain to conduct spatial-spectral fusion [19], but it has two disadvantages. The first one is that it neglects the influence of sensor noises, and the other one is that it fails to utilize the correlation between hyperspectral bands. Vivone et al. [20] provided critical descriptions and extensive comparisons of some of the main state-of-the-art pansharpening methods. Chen et al. [15] compared the advantages and disadvantages of several spatial-temporal fusion models and then proposed [21] a Hierarchical Spatiotemporal Adaptive Fusion Model (HSTAFM) to generate a fine spatial-temporal resolution image, which produces consistently lower biases and performs better than previous models.

Addressing this challenge, in this paper, we developed a novel spatial-spectral fusion model called the Segmented Difference Value method (SEGDV) to generate fine the spatial-spectral resolution image and adopted HSTAFM to conduct spatial and temporal fusion. As both spatial-spectral fusion and spatial-temporal fusion have an identical property of high spatial resolution, we may get the pixel-based information with fine spatial, temporal and spectral resolutions. Here, we present a novel spatial-temporal-spectral fusion framework through spatial-temporal fusion and spatial-spectral fusion and then use the fused information for accurate forest type classification.

2. Materials and Methods

Figure 1 presents a flowchart outlining the methods used in this mapping project, including the data collection and pre-processing, spatial-spectral fusion, spatial-temporal fusion, classification scheme and sampling design, training and validation samples' collection and classification accuracy assessment.

2.1. Study Area

The study area is part of the Gan River nature reserve (Figure 2), which is located in the north of Wuyi mountain, between $25°56'30''$ and $26°07'42''$N, $116°15'01''$ and $116°29'06''$E in the east of Jiangxi province, China, covering an area of 1610.01 km^2. The climate is characterized by a subtropical humid monsoon pattern, high temperature and rain in the summer, warm and humid in the winter. The annual mean temperature is 17.5 °C, and the annual mean precipitation is 2100 mm. The forest area accounts for 95% of the whole area. It mainly includes coniferous, broad-leaved, mixed coniferous and broad-leaved forests and bamboo with high species diversity. Coniferous forests mainly include *Pinus massoniana* and China fir. Broad-leaved forests mainly include *Liquidambar formosana*, *Castanopsis sclerophylla*, *Cinnamomum camphora*, etc. Mixed coniferous and broad-leaved forest is regarded as a succession stage of subtropical pioneer community Masson pine (*Pinus massoniana*) forest being converted into evergreen broad-leaved forest communities [22]. Bamboos include *Phyllostachys heterocycla*, *Bambusa rigida*, etc. China fir, mixed and bamboo forests are the three main forest types in the study area. Bamboos mainly are located in the south of the study area; most China fir is located in the east and a few are located in the west of the study area; mixed forests mainly are located in the middle-west of the study area. *Pinus massoniana*, broad-leaved forest, shrub and farm are less distributed. *Pinus massoniana* mainly is located in the middle-west and east of the study area; broad-leaved forest mainly is located in the southeast, middle and northeast of the study area; fragmentary shrub is located in the southwest and northeast of the study area. The landscape of the study area is mountainous with elevation ranging from 250–1389 m.

All the abbreviations can be found in the list of abbreviations at the end of the manuscript.
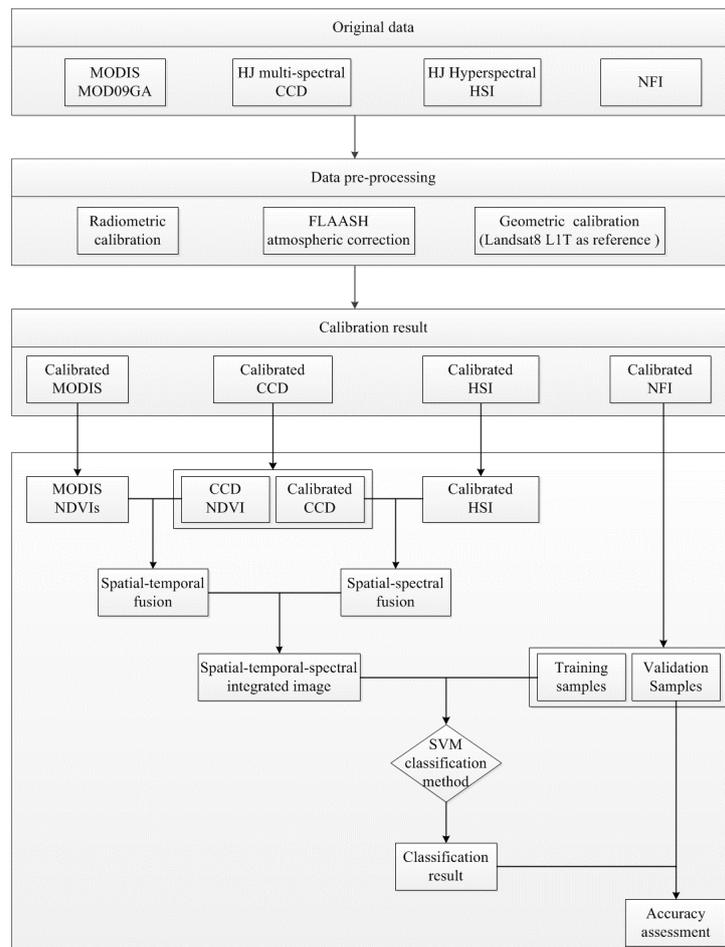
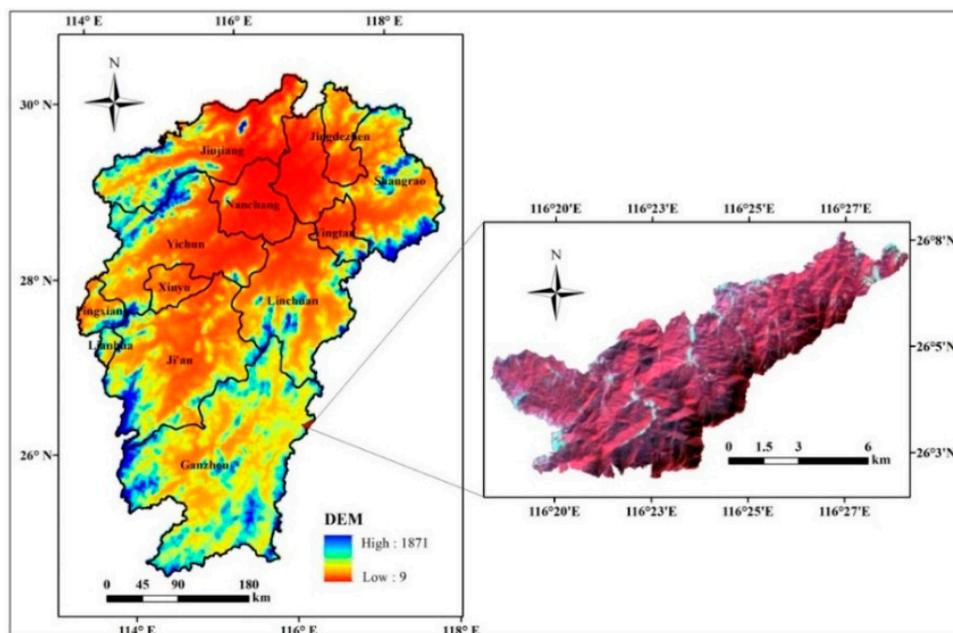**Figure 1.** The flowchart of the research methodology.

Figure 2. The location of the study area in Jiangxi province, DEM of Jiangxi province and the CCD image of the study area.

2.2. Remote Sensing Data

In order to obtain accurate forest information with fine spatial and spectral resolutions, the multi-spectral Charge Coupled Device (CCD) and Hyperspectral Imager (HSI) sensors carried on the China environment (HJ) series satellite, which was launched in September 2008, were used. The CCD contains 4 spectral bands including 0.43–0.52 μm, 0.52–0.60 μm, 0.63–0.69 μm and 0.76–0.90μm, with 30-m spatial resolution per pixel. HSI contains 115 spectral bands ranging from 0.45 μm–0.95 μm, and the band interval is less than 1 nm, with 100-m spatial resolution per pixel. CCD and HSI data were all acquired on 20 October 2012. The data can be downloaded from the website of China center for resources satellite data and application [23]. Spectral radiometric calibration, accurate geometric correction and atmospheric correction were conducted first. To be specific, spectral radiometric calibration was conducted by the ENVI-HJ1A1B-tools. Geometric correction and resampling were conducted in the ENVI 5.1 software using the Landsat 8 L1T images on 5 October 2013 as the space reference basis, so that the CCD and HSI could be consistently matched within the spatial domain. Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) was used to conduct atmospheric correction in the ENVI 5.1 software.

The time series profiles of the vegetation index, such as the Normalized Difference Vegetation Index (NDVI), can describe some important phenological information to monitor the vegetation growth status, determine whether the targeted forest is evergreen or deciduous and estimate the approximate date when the leaves green-up and fall-off. Therefore, the temporal phenological information will be an important factor to differentiate forest types, in addition to the spectral information. In order to obtain a time series dataset of the study area, all the MOD09GA reflectance images with no cloud contaminations in the year 2012 were selected, because only a few CCD images are available because of the serious contaminations of cloud and haze. MOD09GA data were downloaded from the NASA website [24]. All the data were projected to a Universal Transverse Mercator Projection (UTM) Zone 50 N coordinate system using the MODIS reprojection tool (MRT) and then resampled to 30-m spatial resolution via the cubic convolution method in ENVI 5.1 software. A total of 21 cloud-free images was selected in 2012 (Day of Year 51, 87, 88, 94, 270, 271, 278, 284, 286, 287, 293, 294, 295, 307, 311, 318, 341, 344, 348, 360 and 366). Geometric correction and co-registration were conducted in the ENVI 5.1 software using the Landsat images as the base reference, so that all the images could be well matched in the spatial extent. The NDVI was calculated using the MODIS red and near-infrared bands according to Equation (1). Median values in the two nearest days were used to replace the outliers, which are easy to separate, because they present abrupt high or low, compared to ordinary, values. After the median values process, the NDVIs in a time series could be used to better represent forest growth information.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

Landsat has become the longest-running civilian Earth-observing program and the world's largest collection of Earth imagery, since the first satellite was launched in 1972 [25]. It has been used to meet a wide range of information needs due to its 30-m spatial resolution and 16-day revisiting period. Hansen et al. [26] analyzed global 21st-Century forest covers' change based on Landsat data. Lehmann et al. [27] used time series Landsat data for forest cover trends' information of the Australian continent. Zhu et al. [28] developed a new algorithm for Continuous Change Detection and Classification (CCDC) of land cover using all available Landsat data. Meanwhile, it is also widely used in many other applications, such as the estimation of biophysical variables, phenology information, and so on [25]. In this study, Landsat-8 OLI was used as the reference image to verify the effectiveness of the proposed method. All the remote sensing data used in this study and their detailed information can be found in Table 1, and their spectral distributions are provided in Figure 3.

Table 1. The remote sensing data used in the study.

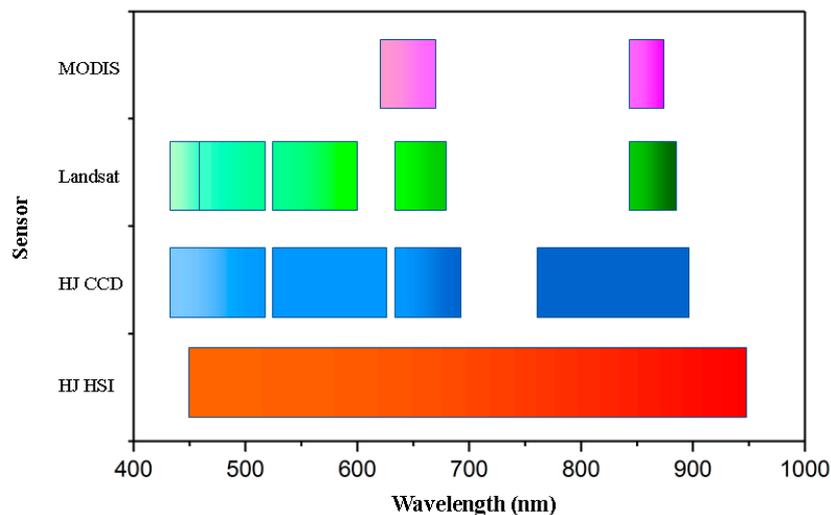|	Spatial Resolution (m)	Revisiting Period (Day)	Band Numbers (Bands)	Usefulness of Data
MOD09GA	250	1	7	Capture phenological information and conduct spatial-temporal fusion
HJ-1A CCD	30	4	4	Capture spatial information and conduct spatial-spectral fusion and spatial-temporal fusion
HJ-1A HSI	100	4	115	Capture spectral information and conduct spatial-spectral fusion
Landsat 8 OLI	30	16	9	Experimental validation with data fusion

Figure 3. The spectral distributions of HSI (115 bands) versus CCD, Landsat 8 and MODIS in the 400–1000-nm range. Each color means one type of sensor, and each block means one band of the corresponding sensor.

2.3. Chinese NFI Data

In order to obtain the area, composition and distribution of forest resources, the State Forestry Bureau of China has already organized eight forest inventories every five years since 1975 [29]. The province-based unit is called the first level forest resource investigation. In each province, the local forestry bureau will carry out the detailed investigation based on the county unit, which is called the second level forest resource investigation. An important component of the investigation is to find out the categories, areas and quality of each forest type in every county, so that the investigation result can objectively reflect the relationship between the forest situation and the factors of local nature, economy and management. Many useful suggestions can be proposed as guides to effectively protect and use forest resources. In the process of investigation, the survey samples were systematically allocated based on statistics theory. Applying unified technical standards of the continuous inventory method, investigators revisited the survey sample sites periodically, processed the data and obtained regional/national forest information using statistical software. According to unified accuracy requirements, the forest stock in each county achieves greater than 80% accuracy at the 95% confidence interval. The tree species and their corresponding proportion of public forests in each county are also 80% accurate at the 95% confidence interval. The positions' accuracy is less than 0.5 mm if there are obvious objects on the ground that can be interpreted as the reference boundary. If not, the positions' accuracy is extended to less than 1 mm.

The data we used in this study come from the second level forest investigation in Jiangxi province. It is a digital forest map providing edge to edge coverage in Jiangxi province, and we set the part of the Ganjiang source nature reserve area as the reference data subset. It contains many detailed field investigation variables such as the serial number of each land parcel, village name, average elevation, terrain slope and orientation, soil properties, tree species and their percent, average tree height, diameter, age, stock volume information, and so on.

2.4. Spatial-Spectral Fusion Method SEGDV

In order to overcome the limitation of the CCD spectral resolution and HSI spatial resolution, spatial-spectral fusion was considered to get fine-resolution images. Generally, the spatial-spectral fusion method includes the fusion of panchromatic and multi-spectral, panchromatic and hyperspectral, multi-spectral and hyperspectral images, morphological information and hyperspectral data, and so on. The fusion of panchromatic and multi-spectral is the most mature. It could be classified into 4 categories: substitution based on Principal Component Analysis (PCA) [30] or Intensity Hue Saturation (IHS) [31]; fusion based on the analysis of multi-resolution [32]; fusion based on model optimization [33]; and fusion based on sparse representation [34]. The fusion methods based on PCA or IHS always caused spectral distortion. The fusion method based on sparse representation has achieved greater success than the component substitution method; but it is commonly very complex, and its efficiency is very low. Fauvel et al. [35] proposed a data fusion scheme for the classification of urban land based on the fusion of the morphological information and hyperspectral data, which succeeded in taking advantage of the spatial and the spectral information simultaneously. The fusion method based on model optimization built the relationship between panchromatic and multi-spectral images and obtained higher fusion accuracy, so the model optimization method was assumed to be the best choice in this article.

The hyperspectral image commonly has high correlation between its bands. Let us denote the digital number of a pixel as $dn_{i,n}$ for band i and pixel n, $I = 1, \ldots, I$; $n = 1, \ldots, N$. I and N are the total number of bands and total number of pixels in an image, respectively [36]. The vector representation for pixel n is $dn_n = (dn_{1,n}, \ldots, dn_{I,n})^{\mathrm{T}}$. The band mean vector is written as $\mu = (\mu_1, \ldots, \mu_i, \ldots, \mu_I)^{\mathrm{T}}$, where μ_i is the mean digital number of band i. The total covariance of the image is represented by:

$$\sum_T = \frac{1}{N} \sum_{n=1}^{N} (dn_n - u)(dn_n - u)^{\mathrm{T}} \tag{2}$$

Figure 4 shows the band correlation matrix of the hyperspectral data. The diagonal line indicates the highest correlation, 1, which is represented in white. The darker the tone, the lower is the absolute value of the correlation. We can see that all the hyperspectral bands are highly correlated except the former noise bands (the dark black part) [36]. Because the main component of these bands is noise and they have little relation with the signal, the value of the correlation between these bands and other bands is very low and appears dark. The contiguous bands along the diagonal line appear "in blocks" showing high correlation among them. The hyperspectral bands were separated into 4 groups according to their correlation with 4 multi-spectral bands.

Then, the fusion between multi-spectral and hyperspectral images can be transformed to the fusion between panchromatic and multi-spectral images in each group. The commonly-used method for multi-spectral and hyperspectral fusion is RIBSR [18,37], but it did not consider the system noise of the sensor. SEGDV was proposed to conduct spatial and spectral fusion to generate the simulated hyperspectral image. For a fixed pixel, the spectral curve profiles of CCD and HSI were discrepant, but they must present a similar change trend, because the reflectance of an object must keep consistent, no matter which sensor was used. Suppose that for the same spectrum range, the reflectance value difference was caused by the systematic errors between the CCD and HSI sensors. We also suppose

that the systematic error is independent of wavelength, which means the error is constant for all the bands. For two random points a, b in spectral curve, we can express this as follows:

$$CCD_a = A + \alpha \tag{3}$$

$$CCD_b = B + \alpha \tag{4}$$

$$HSI_a = A + \beta \tag{5}$$

$$HSI_b = B + \beta \tag{6}$$

CCD_a, CCD_b, HSI_a and HSI_b represent the reflectance value for spectrums *a* and *b* recorded by **CCD** and HSI, respectively. *A*, *B*, *a* and *β* represent the true value and the systematic error of CCD and HSI for spectrums *a* and *b*, respectively. After some operations, we could get:

$$CCD_a - CCD_b = HSI_a - HSI_b \tag{7}$$

However, the spectrum range of CCD is large relative to that of HSI. There are no such bands in the actual multispectral and hyperspectral images that can cover the same spectral wavelength [17]. In this paper, all the HSI bands in the spectrum wavelength range of the CCD were averaged to match the CCD. For example, HSI Bands 26–53 (0.519–0.602 μm) were used to match CCD Band 2 (0.52–0.603 μm); HSI Bands 61–75 (0.632–0.692 μm) were used to match CCD Band 3 (0.63–0.693 μm); HSI Bands 88–110 (0.759–0.909 μm) were used to match CCD Band 4 (0.76–0.903 μm). Because the noise of HSI is very heavy in the first 25 bands (0.46–0.516 μm) and there are some outliers after atmospheric correction in these bands, these bands corresponding to CCD Band 1 (0.43-0.52 μm) were not computed. In this way, the means of each HSI group could be matched with the corresponding CCD band value.
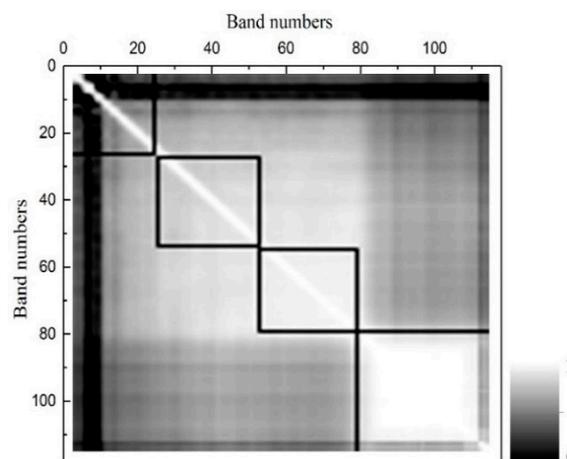

Figure 4. The correlation between hyperspectral bands (the darker the tone, the lower is the absolute value of the correlation between bands.) and their separation into 4 groups according to their correlation with 4 multi-spectral bands.

Most HSI bands have been matched with CCD, but some HSI band spectra still are not covered by CCD, such as HSI Bands 54–60 and 76–87. The hyperspectral images always have a high band correlation. The correlation between CCD bands and HSI bands was calculated. It shows that HSI Bands 26–53 have the highest correlation coefficient with CCD Band 2; HSI Bands 54–81 have the highest correlation coefficient with CCD Band 3; HSI Bands 82–115 have the highest correlation coefficient with CCD Band 4. According to above mentioned knowledge, all the HSI bands were separated into 3 groups. The CCD value and average of HSI in the corresponding group were seen as the basis to generate simulated images with fine spatial and spectral resolution using Equation (7).

2.5. Spatial and Temporal Fusion Model HSTAFM

Spatial-temporal fusion techniques have generated great interest within the remote sensing community, because they can blend multi-spectral and temporal characteristics to generate synthetic data with fine resolutions [21,37–46]. Many advances have been made in the spatiotemporal fusion models, which can be classified into four major categories: (i) transformation-based, (ii) reconstruction-based, (iii) unmixing-based and (iv) learning-based models [15]. Transformation-based models include wavelet and tasseled cap transformations [38,39]. They mainly focus on the integration of spatial and spectral information for image enhancement, instead of constructing a distinct fusion scheme between spatial and temporal information. In the reconstruction-based models, the fusions are generated by a weighted sum of the spectrally-similar neighboring information from fine spatial, but coarse temporal resolution, and fine temporal, but coarse spatial resolution data [40–42]. The unmixing-based models rely on the pixel unmixing techniques, which downscale the coarse resolution images to generate fine-resolution synthetic images while preserving the spectral richness and fidelity [43–45]. In the learning-based models, sparse representation and dictionary learning techniques have generated wide interest [46]. One of the greatest strengths of the learning-based models is that they can predict both phenology and type changes. However, they only use the statistical relationship between the fine and coarse resolution image pair instead of taking any physical properties of remote sensing signals and combining the physical temporal change in the fusion procedure. Although many advances have been made, several shortcomings still remain in existing methods. In order to address the limitations of existing spatiotemporal fusion models and detail time series phenology features of forest to improve classification accuracy, the Hierarchical Spatiotemporal Adaptive Fusion Model (HSTAFM) was proposed [21]. It was used in this study to blend HJ-1A CCD and MODIS images to generate time series fusions with both fine spatial and temporal resolutions. Compared with other spatiotemporal fusion models, the HSTAFM has the following highlights: (i) it combines sparse representation techniques into the physical fusion procedure; (ii) it can predict arbitrary temporal changes including both seasonal phenology change and type change using only one image pair; (iii) it introduces a prior detection of temporal change and a two-level selection strategy of similar pixels, which ensures the accurate capturing of temporal change information.

The implementation of HSTAFM includes two major steps: (i) super-resolution of the coarse-resolution image based on sparse representation; (ii) prediction of the synthetic data by combining the fine-resolution image derived from Step (i). In the first stage, super-resolution of MODIS data was first performed to enhance their spatial resolution by CCD-MODIS image pair dictionary learning. As the transitive-resolution image derived from the first stage is much closer to the actual CCD image in spatial detail, it can be assumed that the pixel purity between the transitive-resolution image and CCD image is approximate. Therefore, the conversion coefficients from the prior/posterior to predicted time between transitive- and fine-resolution (i.e., CCD) images can be assumed to be equal.

$$V_f(x,y,b) = V_t(x,y,b) = \frac{T_2(x,y,b)}{T_1(x,y,b)} \tag{8}$$

where V_f and V_t represent the conversion coefficient of the fine- and transitive-resolution images. T_1 and T_2 denote the transitive-resolution images at prior/posterior time (t_1) and predicted time (t_2). (x,y) is the location of a given pixel, and b denotes the b-th band. After the conversion coefficients have been calculated according to Equation (8), the initially predicted fine-resolution image at t_2 can be obtained through:

$$F_2(x,y,b) = F_1(x,y,b) \cdot V_f(x,y,b) \tag{9}$$

where F_1 denotes the actual fine-resolution image (i.e., CCD) at prior/posterior time (t_1) and F_2 denotes the initially predicted fine-resolution image at the predicted time (t_2). The reflectance difference is computed between the initially predicted fine-resolution image at t_2 and the actual

fine-resolution image at t_1. The difference is employed to explain how much the temporal change is instead of identifying which specific change it is. Here, all possible temporal changes are categorized into two classes: significant change (including land cover change and phenology disturbance) and non-significant change (seasonal phenology change). Each class will be tackled with different strategies to select similar pixels.

After a two-level selection of similar pixels and weight calculation, the final predicted fine-resolution image at the predicted date can be computed through Equation (10). Each prediction of the central pixel's reflectance will be incorporated with spatial and spectral information from its corresponding sets of similar pixels P_{ij}.

$$F(x_{\omega/2}, y_{\omega/2}, b) = \sum_{i=1}^{\omega} \sum_{j=1}^{\omega} W_{ij} \cdot P_{ij} \cdot F_1(x_i, y_j, b) \cdot V_f(x_i, y_j, b) \tag{10}$$

$$W_{ij} = \frac{1/(s_{ij} \cdot d_{ij})}{\sum\limits_{i=1}^{\omega} \sum\limits_{j=1}^{\omega} 1/(s_{ij} \cdot d_{ij})} \tag{11}$$

$$s_{ij} = \frac{P_{ij} \cdot \left| F_1(x_i, y_j) - F_1(x_{\omega/2}, y_{\omega/2}) \right|}{\sum\limits_{i=1}^{\omega} \sum\limits_{j=1}^{\omega} P_{ij} \cdot \left| F_1(x_i, y_j) - F_1(x_{\omega/2}, y_{\omega/2}) \right|} \tag{12}$$

$$d_{ij} = 1 + \sqrt{(x_i - x_{\omega/2})^2 + (y_j - y_{\omega/2})^2}/(1 + \omega) \tag{13}$$

where F denotes the final predicted fine-resolution image. ω denotes the moving window size. (x_i, y_j) and $(x_{\omega/2}, y_{\omega/2})$ denote the locations of candidate similar pixels and central pixels, respectively. P_{ij} is a binary matrix denoting the set of similar pixels, and V_f is the conversion coefficient matrix. W_{ij} is a combined weight determined by the spectral and distance differences [21] according to Equations (11)–(13).

HSTAFM was tested using both the simulated and observed dataset, comparing with the other three state-of-the-art algorithms including the Spatial and Temporal Adaptive Reflectance Fusion model (STARFM), the Flexible Spatiotemporal Data Fusion Model (FSDAF) and the dictionary learning-based Spatiotemporal fusion model using only One base Landsat-MODIS image pair (SP-One) [20], and HSTAFM achieved the highest accuracy. Therefore, this model was used to conduct spatial and temporal fusion in this paper. The HSTAFM algorithm was realized in MATLAB 2016a software. The NDVIs of CCD and MODIS taken on 29 October 2012 were used as the basis images to generate the NDVIs of other 20 days for which MODIS images were required. Through spatial and temporal fusion, we could get the vegetation phenology information with fine spatial and temporal resolution.

2.6. SVM Classification

After spatial-spectral fusion and spatial-temporal fusion, both fused images have identical spatial reference and resolution, so they can be combined together to form a new dataset. After feature combination, a total of 118 variables including 95 spectral variables (Bands 26–110, covering 518–910 nm) and 23 temporal variables are derived, and all the variables are used for classification. We do not perform additional feature reduction before classification since some experiments have already demonstrated that band number reduction or feature extraction (such as PCA and wavelet) of hyperspectral data cannot significantly improve the accuracy compared to just using multispectral data in the classification procedure [36]. In addition, the study area is not very large, so forest type classification can be efficiently performed using all useful variables derived from the fusion.

An SVM classifier was used to map various forest types, because SVM has been proven an effective way to perform hyperspectral classification [47]. In this study, classification and probability estimation

were performed using an SVM classifier with a radial basis function kernel. A brief description of SVM is presented in the following. Assume there are l observations from two classes:

$$(x_1, y_1)(x_2, y_2) \ldots (x_l, y_l) \Big| x_i \in R^N, y_i \in \{-1, 1\} \tag{14}$$

x_i denotes the samples. y_i is a collection of labels that represent the category of x_i. i is the i-th sample. Let us assume that two classes are linearly separable. This means that it is possible to find at least one hyperplane (linear surface) defined by a vector $w \in R^N$ and a bias $b \in R$ that can separate the two classes without errors. Finding the optimal hyperplane involves solving a constrained optimization problem using a quadratic equation. The optimization criterion is the width of the margin between the classes. The discrimination hyper-plane is defined as follows:

$$f(x) = \sum_{i=1}^{l} y_i a_i k(x, x_i) + b \tag{15}$$

where $k(x, x_i)$ is a kernel function and where the sign of $f(x)$ denotes the membership of x. Constructing the optimal hyperplane is equivalent to finding all nonzero a_i values, which are called Lagrange multipliers. Any data point x_i corresponding to a nonzero a_i is a support vector of the optimal hyperplane. A desirable feature of SVMs is that the number of training points that are retained as support vectors is usually quite small, thus rendering them compact classifiers. More information on SVMs is provided in [48,49].

Forest, farm and shrub are identified as the three main land covers types in the study area. The forest can be separated into 5 categories according to the dominant tree species: China fir, *Pinus massoniana*, broad-leaved, mixed and bamboo forest. Because too many broad-leaved tree species co-existed in their distributed area, it is impossible to delimit their boundaries such that all the broad-leaved tree species are incorporated into one category. A total of seven object types, China fir, *Pinus massoniana*, broad-leaved, mixed, bamboo forest, shrub and farm, were identified in the classification. All the Regions Of Interest (ROIs) for training samples and validating accuracy were processed using the ENVI 5.4 software. We used Chinese NFI data and the Statues of Forest Resources Report (http://www.forestry.gov.cn/) to determine the ROIs of forest types at the local level. A total of 387 ROIs were selected, and all the reference ROIs were divided into two groups: 281 for training samples and 106 for evaluating the classification accuracy. SVM was conducted in ENVI 5.4 software. The Radial Basis Function was set as the kernel function of SVM. The Gamma of the kernel function was set to the system default value of 0.034, and the penalty parameter was set to 100.

3. Results

3.1. Results of Spatial and Spectral Fusion

In order to acquire images with fine spatial and hyperspectral resolution, the SEGDV model was used to conduct spatial-spectral fusion between multispectral CCD and hyperspectral HSI images. The algorithm was realized in IDL 8.6 software. Because the CCD and HSI sensors were carried on the same platform and the images were both taken on 20 December 2012, they have identical spatial reference and similar spectral information and could be matched together in the spatial and spectral domain. Figure 5 displays the standard false color composite image (NIR-R-G) of the original CCD, HSI and SEGDV fusion image. We could find that the SEGDV fused image could retain the detailed spatial resolution from the CCD data while preserving the consistent spectral information of the original HSI data. We further randomly selected one sample pixel in the homogeneous area of each category. Figure 6 shows the multi-spectral profiles and fused hyperspectral profiles of the selected sample. It could be found that the fused image greatly enhanced the spectral resolution compared with the multispectral CCD image, and the spectral profiles of CCD and the corresponding profiles of

HSI have kept the same walking trend, which demonstrated the effectiveness of SEGDV in blending spatial-spectral information.
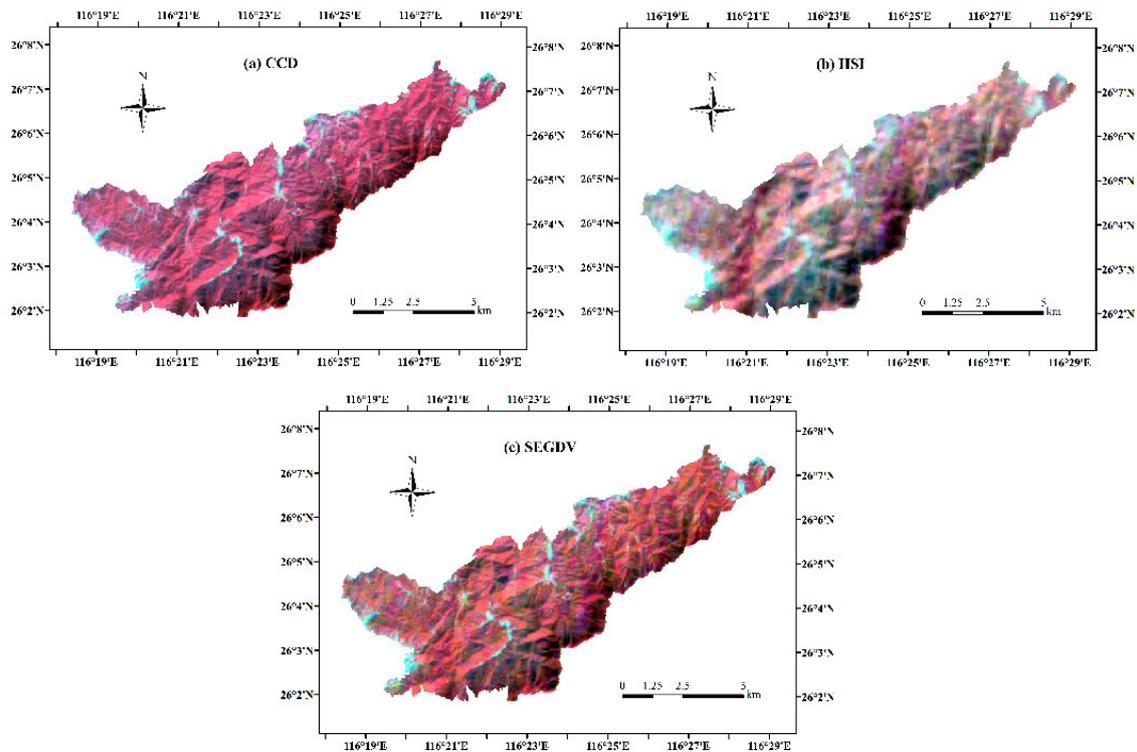

Figure 5. The visual effect of the original CCD, HSI and fused image. (**a**) Original CCD (Bands 4, 3, 2 composite); (**b**) original HSI (Bands 100, 66, 47 composite); (**c**) fused image (Bands 100, 66, 47 composite).
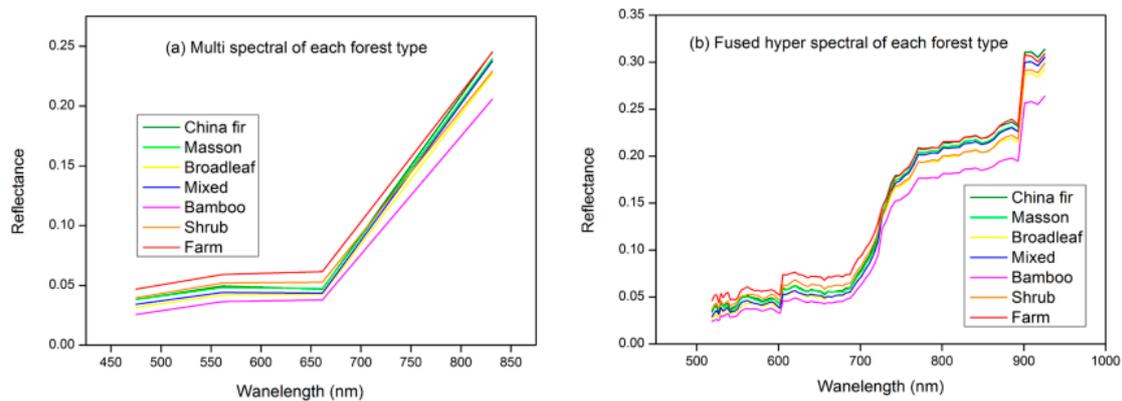

Figure 6. The multiple and fused hyperspectral curves of different forest types. The multi spectral curves (**a**) and the fused hyperspectral curves (**b**) of different forest types.

3.2. Results of Spatial and Temporal Fusion

In order to obtain detailed forest phenology information, which is an important variable to describe and distinguish different forest types, the HSTAFM model was used to conduct spatial-temporal fusion between CCD and MODIS images [21]. The HSTAFM algorithm was realized in MATLAB 2016a software. The CCD and MODIS images taken on 20 December 2012 and another MODIS image taken on the predicted date were used as the basis to predict the unknown CCD on the predicted date. As shown in Figure 7, we could find that the fused image could enhance the spatial

information significantly compared with the original MODIS image. The seven above-mentioned sample pixels were used again to describe the time series changes of various forest types. In Figure 8, it could be found out that most forest spectral profiles are well in accordance with the actual forest phenology. The forest began to grow in spring, and the NDVI increased in February–April. The NDVI reached its peak in September and began to decrease in December. Differences among the NDVI time series profiles of different forest types are very important to differentiate different forest types.
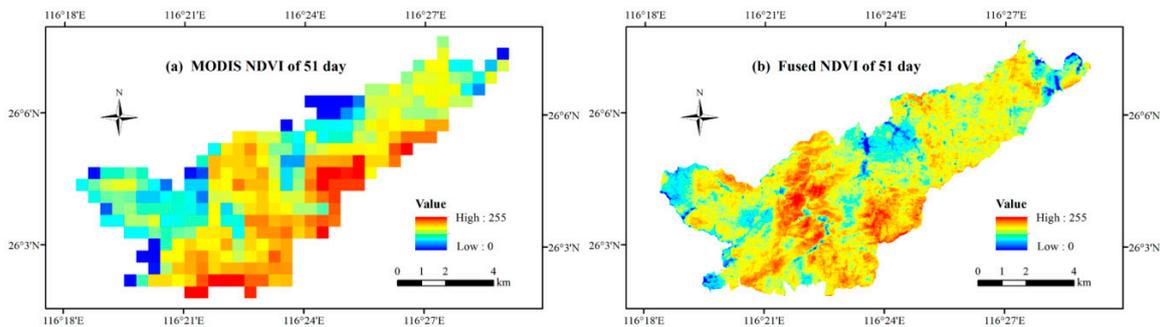

Figure 7. The NDVI of (**a**) MODIS and (**b**) after spatial and temporal fusion on 20 February 2012.
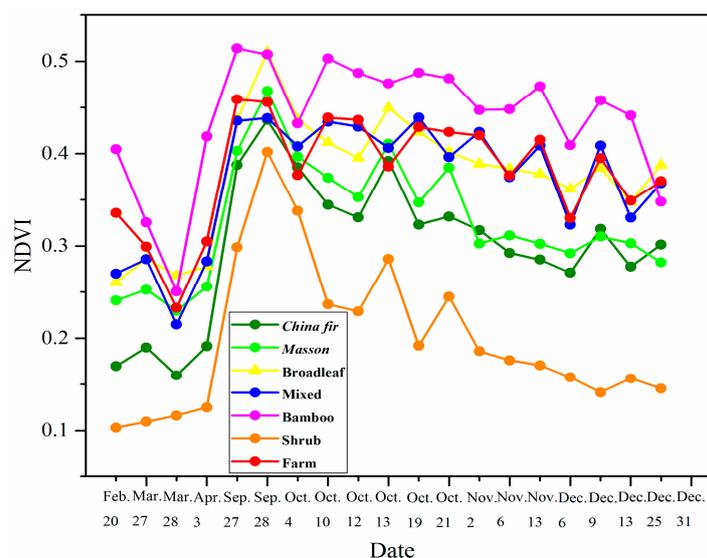

Figure 8. The NDVI time series curves of China fir, *Pinus massoniana*, broad-leaved, mixed, bamboo, shrub forests and farm in 2012.

3.3. Forest Type Classification

The classification results are evaluated to determine the accuracy and reliability levels. The confusion matrix is used to compare the number of pixels divided into a class by the number of pixels for the class in the ground truth (Figure 9). Every column denotes the ground truth. Each line shows the classification results. As the value increases along the diagonal, the level of accuracy increases. The overall accuracy level is the ratio of correctly classified pixel numbers to all pixel numbers. The commission error refers to pixels categorized into a class of interest that belong to other classes. The omission error refers to pixels that belong to the real classification of the surface, but that are not correctly classified by the classifier. The producer's accuracy level is the ratio of pixels correctly classified into a class of interest to the total number of ground truth pixels of the class of interest. The user's accuracy level refers to the ratio of pixels correctly classified under the class of interest to the number of total pixels classified under the class of interest by a classifier [50].

Actual Class

		A	B	C	D	Σ
	A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
	B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
Predicted Class	C	n_{CA}	n_{CB}	n_{CC}	n_{CD}	n_{C+}
	D	n_{DA}	n_{DB}	n_{DC}	n_{DD}	n_{D+}
	Σ	n_{+A}	n_{+B}	n_{+C}	n_{-D}	n

Figure 9. The confusion matrix of classification. The highlighted elements represent the main diagonal of the matrix that contains the cases where the class labels depicted in the image classification and ground dataset agree, whereas the off-diagonal elements contain those cases where there is a disagreement in the labels.

In order to validate the effectiveness of the proposed method, comparison experiments were conducted. Classification results derived from single Landsat 8, single spatial-spectral fusions and spatial-temporal fusions were used to compare with that derived from the spatial-spectral-temporal integrated fusion image. As shown in Figure 10, compared with the NFI data (Figure 10e), which served as the true distribution of various forest types, we could find out that the Landsat-based classification (Figure 10a) failed to separate those less distributed categories, such as broad-leaved, shrub. In addition, the China fir coverage was overestimated, and the mixed forest was underestimated. Although most forest types were correctly classified in the results derived from the spatial-spectral fusion image (Figure 10b), the classification map is too fragmented. As for the classification results derived from the spatial-temporal fusion image (Figure 10c), the majority of forest types were correctly classified, and the fragment phenomenon was also improved significantly, but there still existed some obvious errors, for example the broad-leaved forest was overestimated in the middle-east, while it was underestimated in the southeast of the study area. Generally, the classification derived from the spatial-spectral-temporal fusion achieved the most plausible result (Figure 10d). We could find that most forest types were correctly classified except for some local areas for which a small portion of China fir was still misclassified into broad-leaved forest and some *Pinus massoniana* was underestimated.

From Tables 2 and 3, we can find that most forest types have achieved ideal accuracy except *Pinus massoniana* and broad-leaved forest. It should be pointed out that all the images were classified using exactly the same samples, and the proposed method has achieved the highest accuracy and Kappa coefficient (Table 4). Another interesting point is that the accuracy of spatial-temporal fusion was better than spatial-spectral fusion, showing that the time series phenology information is more effective than spectra information in the classification of different forest types.

Table 2. The confusion matrix of spatial-spectral-temporal integrated fusion.

Class	Bamboo	Farm	Shrub	Broad-Leaved	Masson	China Fir	Mixed	Total
Unclassified	0	0	0	0	0	0	0	0
Bamboo	1076	0	0	38	0	0	39	1153
Farm	0	45	0	0	13	5	0	63
Shrub	0	0	53	0	2	0	0	55
Broad-leaved	36	0	10	144	1	71	5	267
Masson	15	3	3	0	63	19	11	114
China fir	82	0	3	21	17	751	54	928
Mixed	16	0	7	21	24	51	758	877
Total	1225	48	76	224	120	897	867	3457

Table 3. The commission error, omission error, producer accuracy and user accuracy of each forest type.

	Commission (%)	Omission (%)	Prod.Acc (%)	User.Acc (%)
Bamboo	6.68	12.16	87.84	93.32
Farm	28.57	6.25	93.75	71.43
Shrub	3.64	30.26	69.74	96.36
Broad-leaved	46.07	35.71	64.29	53.93
Masson	44.74	47.5	52.5	55.26
China fir	19.07	16.28	83.72	80.93
Mixed	13.57	12.57	87.43	86.43

Table 4. The overall accuracy and Kappa coefficient of each method.

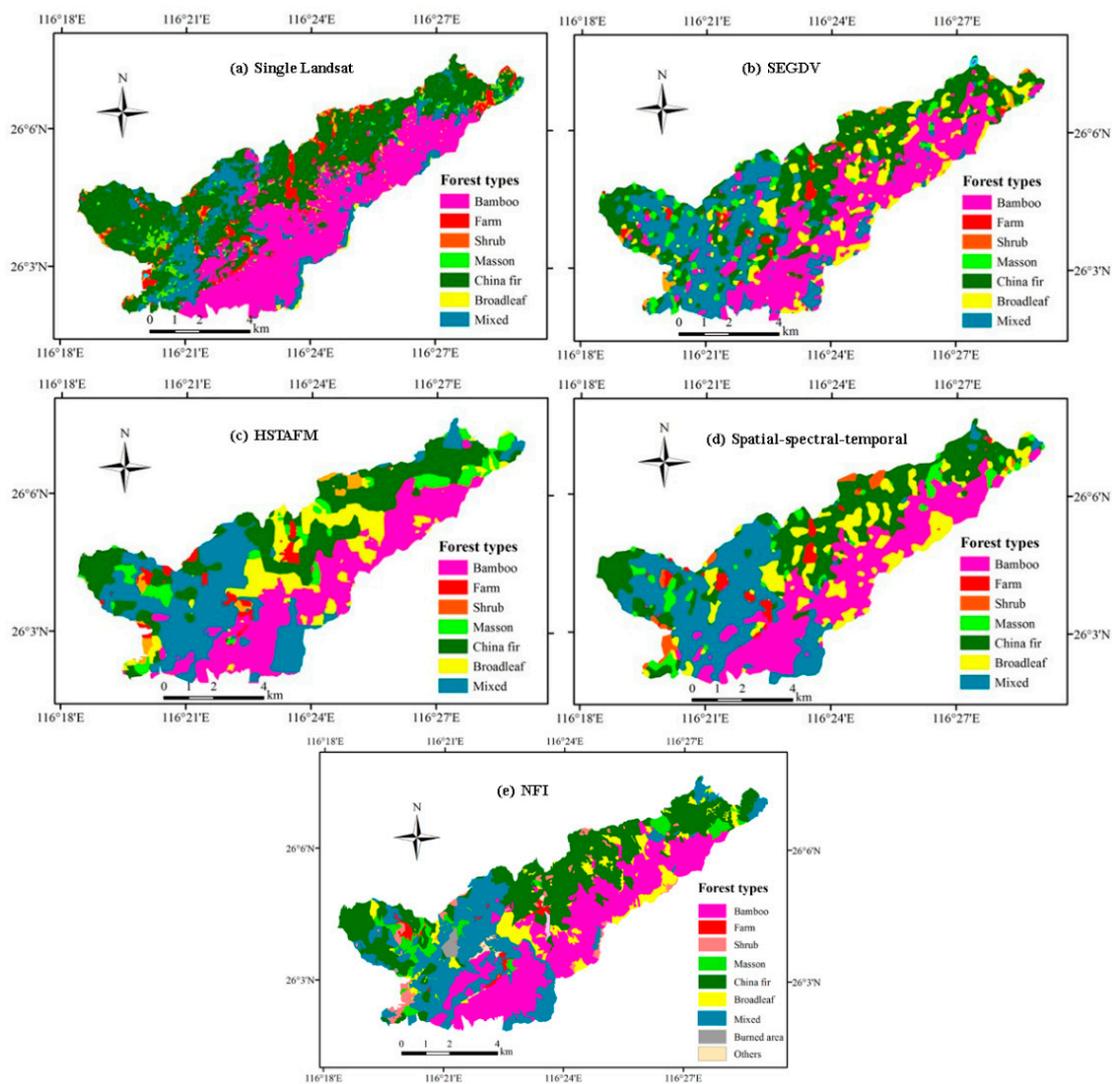|	Landsat Image	Spatial-Spectral Fusion	Spatial-Temporal Fusion	Spatial-Spectral-Temporal Integrated Fusion
Overall accuracy	69.95%	70.95%	78.94%	83.60%
Kappa	0.59	0.61	0.72	0.78

Figure 10. The classification result of: (**a**) single Landsat; (**b**) spatial and spectral fusion, SEGDV; (**c**) spatial and temporal fusion, HSTAFM; (**d**) spatial-spectral-temporal integrated fusion; (**e**) the NFI results.

4. Discussion

4.1. The Classification Result

From Table 4, we can see that the proposed spatial-spectral-temporal fusion has achieved the highest overall accuracy of 83.60%. In Tables 2 and 3, we could also find that most categories have achieved plausible accuracies. However, broad-leaved forest and *Pinus massoniana* did not achieve satisfactory accuracies, which could be caused by two potential reasons. On the one hand, as shown in Figure 6, we could find that the differences of spectral profiles between China fir and *Pinus massoniana* and between broad-leaved forest and shrub curves are very small. Moreover, as shown in Figure 8, the temporal changes of China fir and *Pinus massoniana* also have very similar paces, as was the case for the temporal changes of mixed forest and farm. This inevitably added difficulties to distinguishing these forest types. On the other hand, the NFI data were assumed to be the ground truth of forest types in this study, but in fact, they may still have some biases. In Figure 10e, we could find that the areas of *Pinus massoniana*, farm, shrub and broad-leaved forest are very small so that less samples would be picked up (Table 2). Furthermore, these forest types were sparsely and fragmentarily distributed, because the geographic environment of the Gan River nature reserve was very complex and had rich biodiversity. As their proportions and borders were not clearly differentiated, the existing mixed pixels would further add difficulties in separating them accurately. It is indeed a difficult task to differentiate forest types with a 30-m spatial resolution. A promising approach to solve this problem is to seek new Earth observation data with higher spatial, temporal and spectral resolution to mitigate the effects of mixed pixels. For example, Ren et al. [51] conducted forest land type precise classification based on SPOT-5 and GF-1 images and achieved an overall accuracy of 92.28% with a Kappa coefficient of 0.8996. Unmanned Aerial Vehicles (UAV) might be another effective way to get higher resolution images. Currently, the spatial resolution of the images acquired from UAV observations can even reach 0.05 m, which will lend great support to improving classification accuracy. On the other hand, how to select optimum features, such as the vegetation index, the statistical index or making full use of spatial, spectral and temporal information to separate different forest types, is still a pivotal task.

4.2. The Spatial-Spectral Fusion

In order to demonstrate the advantage of SEGDV, a control experiment was conducted. We should be aware of the fact that no actual hyperspectral image with high spatial resolution existed in our study area. An emerging problem is how to quantitatively validate the fusion result without no reference data. Here, we resampled the HSI with 100-m spatial resolution to 300-m as a coarse resolution image. We resampled the multispectral CCD with 30-m spatial resolution to 100-m as a fine resolution image. By fusing the up-sample 300-m hyperspectral image and 100-m multispectral image via the proposed SEGDV method, we could obtain the fused 100-m hyperspectral image. In this way, the original un-resampled HSI could be used as the actual reference data to evaluate the fusion accuracy quantitatively. The commonly-used RIBSR method was used for the comparison experiment.

As shown in Figure 11, we could find that the result of SEGDV provided more spatial detail than that of RIBSR, such as the bluish farm in the red circle. Additionally, the SEGDV model also achieved a better visual effect than RIBSR. Two typical land cover types, forest and farm, were selected from the observed and fused data to further investigate the hyperspectral fidelity. As shown in Figure 12, no matter the forest or farm, the reflectance profiles share similar trends between the observed and fused hyperspectral images (Figure 12a,b). We also found that the profile of SEGDV is closer to the real observed *I* data in most spectral ranges, especially for forest. The Root Mean Square Error (RMSE) was adopted to measure the fusion biases between the observed and fused Hyperion data quantitatively. The RMSE is calculated as follow:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{fused,i} - X_{obs,i}\right)^2}{n}} \tag{16}$$

where *n* means the total number of bands, *i* means the *i*-th band, $X_{fused,i}$ the means the reflectance value of the *i*-th band in the fused image and $X_{obs,i}$ means the reflectance value of the observed hyperspectral image. From Figure 13, we can see that in most bands, the RMSE of SEGDV is smaller than that of RIBSR, especially after 800 nm. Above all, we can make the conclusion that SEGDV performs better than RIBSR.
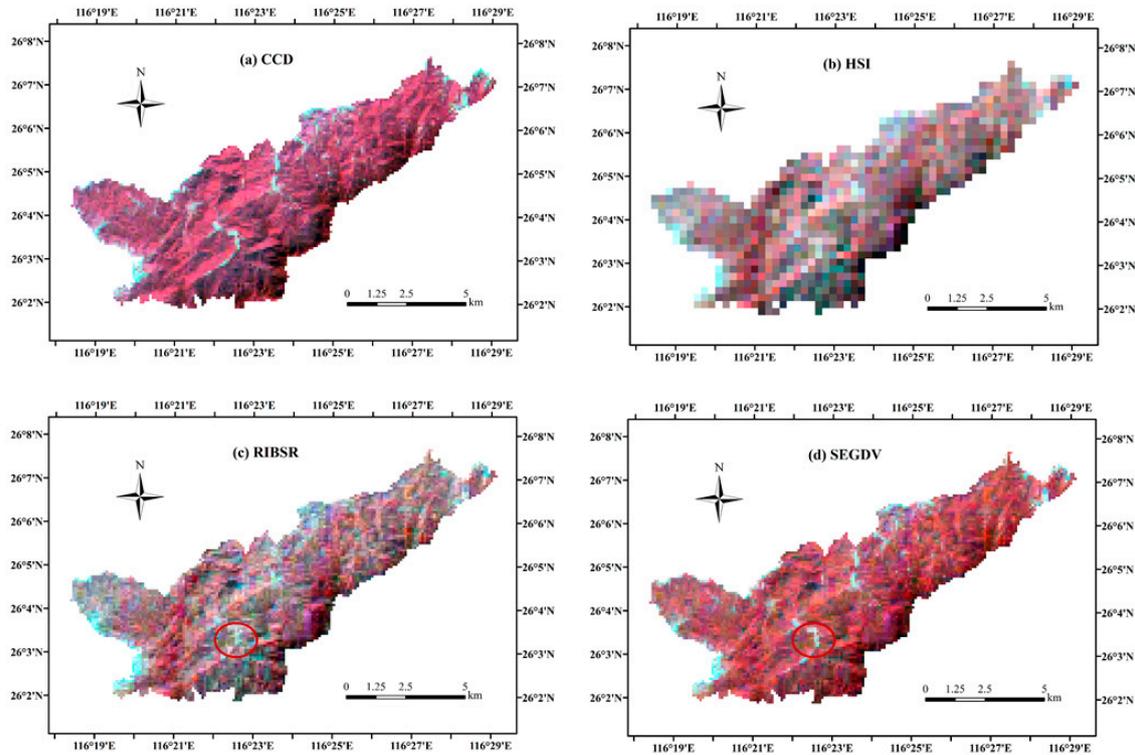

Figure 11. The visual effect of (**a**) resampled CCD 100 m; (**b**) resampled HSI 300 m; (**c**) the RIBSR 100 m and (**d**) SEGDV 300 m fused images.
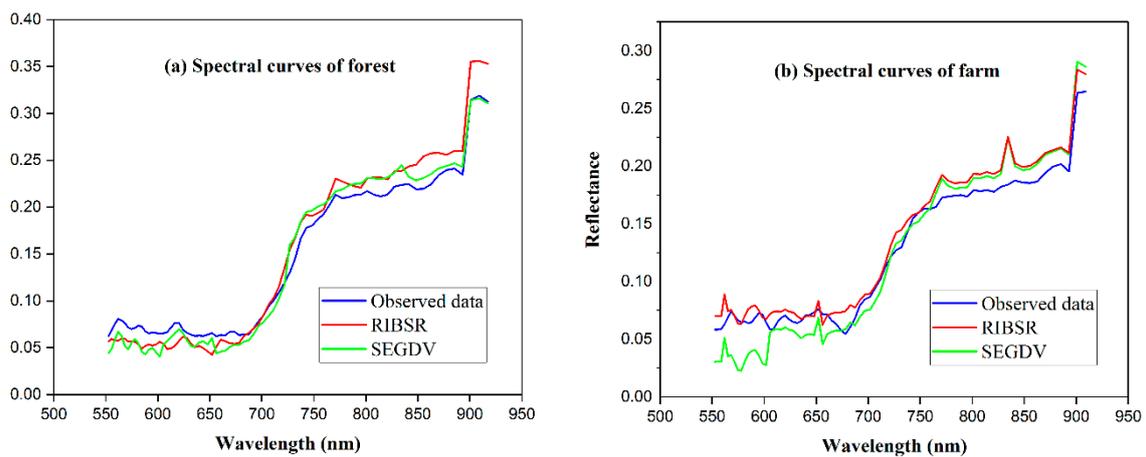

Figure 12. The spectral curves of forest and farm derived from the observed HSI data, RIBSR and SEGDV fusion method.
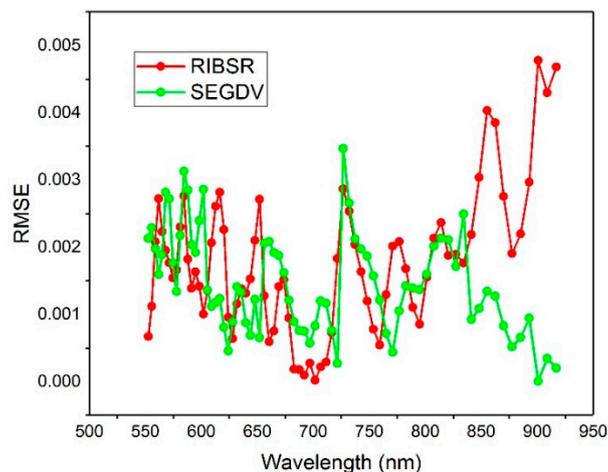
Figure 13. The RMSE of the fused hyperspectral images derived from forest.

Two reasons may contribute to this result. Firstly, SEGDV considered the influence of noise. The commonly-used RIBSR model [18,36] supposed that if I_a and I_b are two images with the same size, then their ratio image I_R can be calculated as follows:

$$I_R(x,y) = I_a(x,y)/I_b(x,y)$$
$$x = 0,\ldots,cols-1;$$
$$y = 0,\ldots,rows-1$$

(17)

x, y means the location of each pixel. Suppose that I_1 and I_2 are two sets of multispectral/hyperspectral images of the same scene and having the same size. $I_1[m_1]$ and $I_1[m_2]$, $I_2[n_1]$ and $I_2[n_2]$ are two different bands in I_1 and I_2, respectively. $I_1[m_1]$ and $I_2[n_1]$ have the same spectral range, so do $I_1[m_2]$ and $I_2[n_2]$. According to the fact that the ratio of reflectivity of the same kind of land cover in two given spectral ranges is almost changeless, their relationship can be depicted as follows:

$$\frac{I_1[m_1]}{I_1[m_2]} = \frac{I_2[n_1]}{I_2[n_2]}$$

(18)

However, the hyperspectral images are commonly affected by noise [52–54], and when noise exists, their relationship will not be like this. Secondly, SEGDV considered the relationship between different bands. SEGDV separated the whole hyperspectral range into several groups. In each group, different bands were highly correlated with each other, which reduced the errors in spatial and spectral fusion. Of course, this model also has its problems. In this model, we assume that a band in a multispectral image corresponds to several bands in a hyperspectral image with a narrower wavelength range, but this may be not the case. The relationship of the spectra range between HSI bands and CCD still needs to be considered further.

4.3. The Spatial-Temporal Fusion

Although the spatial resolution has been enhanced greatly after HSTAFM fusion, at the same time, some errors caused by mixed pixels were still brought in, which eventually led to the uncertainty of the classification results. The low spatial resolution MODIS images commonly cannot capture the spatial difference in small areas due to its large spatial resolution (500 m), such that the mosaic phenomenon is very obvious (Figure 7a). A MODIS pixel (250 × 250 m) was unmixed into about 64 pixels corresponding to CCD pixels (30 × 30 m), and the NDVI calculated from MODIS images became the basis to conduct spatial-temporal fusion. The mixed pixel and mosaic problems were eventually transmitted to the fusion results as shown in Figure 7b. Some pixels belong to the same

one category, but their value showed a big difference. Meanwhile, there are also some outliers in the time series NDVI profiles (Figure 8), such as the NDVI values on 6 and 9 December, which may be caused by subpixel clouds, variable illumination conditions and viewing geometries and other remnant geometric errors. Therefore, how to improve the quality of the data source will be an important step to produce the final classification results with high accuracy.

5. Conclusions

In this article, we proposed a spatial-spectral-temporal fusion framework through the SEGDV spatial-spectral fusion model and the HSTAFM spatial-temporal fusion model. The fused image with fine resolution was used to classify different forest types. The entire research method could be divided into five parts. First is the data preprocess including the projection transformation, atmospheric correction and geometric correction to ensure all the images could be well matched in spatial, spectral and temporal domains. Second, multi-source data fusion was conducted including spatial-spectral fusion and spatial-temporal fusion. Third, the fused hyperspectral and multi-temporal images were combined together to form the synthetic fusions, which contained all the spectral and temporal information. Fourth, training and validation samples were selected, and a SVM classifier was used to classify different forest types. Last, the classification result was derived, and the accuracy was estimated. Experimental results showed that compared with the classifications derived from single Landsat-8 image (69.95%), single spatial-spectral fusions (70.95%) and single spatial-temporal fusion (78.94%), the proposed method achieved the highest accuracy of 83.6%, thereby providing a new approach to sub-species classification such as the differentiation of different types of forest, grassland, crop, wetland, and so on.

Author Contributions: Bing Xu and Tianxiang Yue conceived of and designed the experiments. Ming Lu performed the experiments. Xiaohan Liao, Bin Chen and Shengming Ren analyzed the data. Huanyin Yue, Xiaowen Li and Zhen Nie contributed analysis tools. Ming Lu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ALS	Airborne Laser Scanning
CCD	Charge Coupled Device
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
FSDAF	Flexible Spatiotemporal Data Fusion model
HSI	Hyperspectral Imager
HSTAFM	hierarchical Spatiotemporal Adaptive Fusion Model
IHS	Intensity Hue Saturation
LIDAR	Light Detection and Ranging
NDVI	Normalized Difference Vegetation Index
NFI	National Forest resources Inventory
PCA	Principal Component Analysis
RIBSR	Ratio Image-Based Spectral Resampling
ROI	Regions of Interest
SEGDV	Segmented Difference Value
SP-One	Dictionary learning-based Spatiotemporal fusion model using only One base Landsat-MODIS image pair
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
SVM	Support Vector Machine
VLDI	Vegetation Local Difference Index
UTM	Universal Transverse Mercator Projection

References

1. Pan, Y.; Birdsey, R.A.; Phillips, O.L.; Jackson, R.B. The structure, distribution, and biomass of the world's forests. *Annu. Rev. Ecol. Evol. Syst.* **2013**, *44*, 593–622. [CrossRef]

2. Masek, J.G.; Hayes, D.J.; Hughes, M.J.; Healey, S.P.; Turner, D.P. The role of remote sensing in process-scaling studies of managed forest ecosystems. *For. Ecol. Manag.* **2015**, *355*, 109–123. [CrossRef]

3. McKinley, D.C.; Ryan, M.G.; Birdsey, R.A.; Giardina, C.P.; Harmon, M.E.; Heath, L.S.; Houghton, R.A.; Jackson, R.B.; Morrison, J.F.; Murray, B.C. A synthesis of current knowledge on forests and carbon storage in the United States. *Ecol. Appl.* **2011**, *21*, 1902–1924. [CrossRef] [PubMed]

4. Young, R.A. *Introduction to Forest Science*; John Wiley & Sons: Hoboken, NJ, USA, 1982.

5. Qin, Y.; Xiao, X.; Dong, J.; Zhang, G.; Roy, P.S.; Joshi, P.K.; Gilani, H.; Murthy, M.S.R.; Jin, C.; Wang, J. Mapping forests in monsoon Asia with ALOS PALSAR 50-m mosaic images and MODIS imagery in 2010. *Sci. Rep.* **2016**, *6*. [CrossRef] [PubMed]

6. Molinier, M.; López-Sánchez, C.A.; Toivanen, T.; Korpela, I.; Corral-Rivas, I.J.; Tergujeff, R.; Häme, T. Relasphone—Mobile and Participative In Situ Forest Biomass Measurements Supporting Satellite Image Mapping. *Remote Sens.* **2016**, *8*, 869. [CrossRef]

7. Pratihast, A.K.; DeVries, B.; Avitabile, V.; Bruin, S.; Kooistra, L.; Tekle, M.; Herold, M. Combining Satellite Data and Community-Based Observations for Forest Monitoring. *Forests* **2014**, *5*, 2464–2489. [CrossRef]

8. Wulder, M.; Franklin, S. *Remote Sensing of Forest Environments: Concepts and Case Studies*; Springer Science & Business Media: Berlin, Germany, 2012.

9. Ren, C.; Ju, H.; Zhang, H.; Huang, J.; Zheng, Y. Multi-source data for forest land type precise classification. *Sci. Silvae Sin.* **2016**, *52*, 54–65.

10. Torresan, C.; Corona, P.; Scrinzi, G.; Marsal, J.V. Using classification trees to predict forest structure types from lidar data. *Ann. For. Res.* **2016**, *59*, 281–298. [CrossRef]

11. Gorgens, E.B.; Soares, C.P.B.; Nunes, M.H.; Rodriguez, L.C.E. Characterization of Brazilian forest types utilizing canopy height profiles derived from airborne laser scanning. *Appl. Veg. Sci.* **2016**, *19*, 518–527. [CrossRef]

12. Chen, C.; Bian, Z.; Li, S.; Tang, P.; Wu, H. Improving forest type classification using the vegetation local difference index. *Int. J. Remote Sens.* **2015**, *36*, 3701–3713. [CrossRef]

13. Castilla, G.; Rodrigue, S.; Skakun, R.S.; Hall, R.J. Four national maps of broad forest type provide inconsistent answers to the question of what burns in Canada. *Remote Sens.* **2016**, *8*, 539. [CrossRef]

14. Connette, G.; Oswald, P.; Songer, M.; Leimgruber, P. Mapping distinct forest types improves overall forest identification based on multi-spectral Landsat imagery for Myanmar's Tanintharyi region. *Remote Sens.* **2016**, *8*, 882. [CrossRef]

15. Chen, B.; Huang, B.; Xu, B. Comparison of spatiotemporal fusion models: A review. *Remote Sens.* **2015**, *7*, 1798–1835. [CrossRef]

16. Zhang, L.P.; Shen, H.F. Progress and future of remote sensing data fusion. *J. Remote Sens.* **2016**, *20*, 1050–1061.

17. Shen, H. Integrated fusion method for multiple temporal-spatial-spectral images. In Proceedings of the XXII ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012; pp. 407–410.

18. Zhang, Y.F.; He, M.Y. Multi-spectral and hyperspectral image fusion using 3-D wavelet transform. *J. Electron.* **2007**, *24*, 208–244. [CrossRef]

19. Chen, Z.; Pu, H.Y.; Wang, B. Fusion of Hyperspectral and Multispectral Images: A Novel Framework Based on Generalization of Pan-Sharpening Methods. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1418–1422. [CrossRef]

20. Vivone, G.; Alparone, L.; Chanussot, J.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [CrossRef]

21. Chen, B.; Huang, B.; Xu, B. A hierarchical spatiotemporal adaptive fusion model using one image pair. *Int. J. Digit. Earth* **2016**, *10*, 1–17. [CrossRef]

22. Wang, F.; Hu, X.; Liu, Y.; Zhou, F.; Zhou, H.; Liu, Z. Community composition and spatial distribution of coniferous-broadleaved mixed forest in subtropical region, Jiangxi province. *Acta Agric. Univ. Jiangxiensis* **2013**, *35*, 761–768.

23. China Center for Resources Satellite and Application. Available online: http://www.cresda.com/CN/ (accessed on 2 August 2017).

24. LAADS DAAC. Available online: https://ladsweb.modaps.eosdis.nasa.gov/ (accessed on 2 August 2017).

25. Wulder, M.A.; Coops, N.C. Make Earth observations open access. *Nature* **2014**, *513*, 30–31. [CrossRef] [PubMed]

26. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef] [PubMed]

27. Lehmann, E.A.; Wallace, J.F.; Caccetta, P.A.; Furby, S.L.; Zdunic, K. Forest cover trends from time series Landsat data for the Australian continent. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 453–462. [CrossRef]

28. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]

29. Bureau, S.F. Brief Introduction of Forest Resources in China—The Eighth National Forestry Inventory. Available online: http://211.167.243.162:8085/8/chengguobaogao/showpageinit?lm=xxxz (accessed on 2 August 2017).

30. Chavez, P.S.; Sides, S.C.; Anderson, J.A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic, Photogram. *Eng. Remote Sens.* **1991**, *57*, 265–303.

31. Haydn, R.; Dalke, G.W.; Henkel, J.; Bare, J.E. Application of the IHS color transform to the processing of multi sensor data and image enhancement. In Proceedings of the International Symposium on Remote Sensing of Environment, First Thematic Confrence: "Remote Sensing of Arid and Semi-Arid Lands", Cairo, Egypt, 19–25 January1982; pp. 599–616.

32. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on over sampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2300–2312. [CrossRef]

33. Zhang, L.P.; Shen, H.F.; Gong, W.; Zhang, H.Y. Adjustable model-based fusion method for multispectral and panchromatic images. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 1693–1704. [CrossRef] [PubMed]

34. Jiang, C.; Zhang, H.Y.; Shen, H.F.; Zhang, L.P. A practical compressed sensing-based pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 629–633. [CrossRef]

35. Fauvel, M.; Chanussot, J. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3804–3814. [CrossRef]

36. Xu, B.; Gong, P. Land-use/land-cover classification with multispectral and hyperspectral E0–1 data. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 955–965. [CrossRef]

37. Chen, B.; Chen, L.; Lu, M.; Xu, B. Wetland mapping by fusing fine spatial and hyperspectral resolution images. *Ecol. Model.* **2017**, *353*, 95–106. [CrossRef]

38. Acerbi-Junior, F.W.; Clevers, J.G.P.W.; Schaepman, M.E. The assessment of multi-sensor image fusion using wavelet transforms for mapping the Brazilian Savanna. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 278–288. [CrossRef]

39. Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sens. Environ.* **2009**, *113*, 1613–1627. [CrossRef]

40. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218.

41. Zhu, X.L.; Chen, J.; Gao, F.; Chen, X.H.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [CrossRef]

42. Michishita, R.; Chen, L.F.; Chen, J.; Zhu, X.L.; Xu, B. Spatiotemporal reflectance blending in a wetland environment. *Int. J. Digit. Earth* **2015**, *8*, 364–382. [CrossRef]

43. Zhukov, B.; Oertel, D.; Lanzl, F.; Reinhackel, G. Unmixing-based multi sensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [CrossRef]

44. Zurita-Milla, R.; Clevers, J.G.P.W.; Schaepman, M.E. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 453–457. [CrossRef]

45. Zhu, X.L.; Helmer, E.H.; Gao, F.; Liu, D.S.; Chen, J.; Lefsky, M.A. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [CrossRef]

46. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [CrossRef]

47. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]

48. Burges, C.J. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [CrossRef]

49. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]

50. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]

51. Ren, C.; Ju, H.; Zhang, H.; Huang, J. Forest land type precise classification based on SPOT5 and GF-1 images. In Proceedings of the 2016 IEEE International Geoscience and Remote sensing Symposium, Beijing, China, 10–15 July 2016; pp. 894–897.

52. Xu, B.; Gong, P. Noise estimation in a noise-adjusted principal component transformation and hyperspectral image restoration. *Can. J. Remote Sens.* **2008**, *34*, 271–286.

53. Acito, N.; Diani, M.; Corsini, G. Subspace-based striping noise reduction in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1325–1342. [CrossRef]

54. Lu, M.; Hu, L.J.; Yue, T.X.; Chen, Z.Y.; Chen, B.; Lu, X.Q.; Xu, B. Penalized Linear Discriminant Analysis of Hyperspectral Imagery for Noise Removal. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 359–363. [CrossRef]