Remote Sens. **2015**, *7*, 4424-4441; doi:10.3390/rs70404424

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Advancing of Land Surface Temperature Retrieval Using Extreme Learning Machine and Spatio-Temporal Adaptive Data Fusion Algorithm

Yang Bai [1,2], Man Sing Wong [1,*], Wen-Zhong Shi [1], Li-Xin Wu [2] and Kai Qin [2]

[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong; E-Mails: baiyang_cumt@163.com (Y.B.); john.wz.shi@polyu.edu.hk (W.S.)

[2] School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China; E-Mails: awulixin@263.net (L.W.); qinkai20071014@163.com (K.Q.)

* Author to whom correspondence should be addressed; E-Mail: lswong@polyu.edu.hk; Tel.: +852-3400-8959.

Academic Editors: Zhao-Liang Li, Jose A. Sobrino, Xiaoning Song, George P. Petropoulos and Prasad S. Thenkabail

Received: 3 November 2014 / Accepted: 8 April 2015 / Published: 14 April 2015

Abstract: As a critical variable to characterize the biophysical processes in ecological environment, and as a key indicator in the surface energy balance, evapotranspiration and urban heat islands, Land Surface Temperature (LST) retrieved from Thermal Infra-Red (TIR) images at both high temporal and spatial resolution is in urgent need. However, due to the limitations of the existing satellite sensors, there is no earth observation which can obtain TIR at detailed spatial- and temporal-resolution simultaneously. Thus, several attempts of image fusion by blending the TIR data from high temporal resolution sensor with data from high spatial resolution sensor have been studied. This paper presents a novel data fusion method by integrating image fusion and spatio-temporal fusion techniques, for deriving LST datasets at 30 m spatial resolution from daily MODIS image and Landsat ETM+ images. The Landsat ETM+ TIR data were firstly enhanced based on extreme learning machine (ELM) algorithm using neural network regression model, from 60 m to 30 m resolution. Then, the MODIS LST and enhanced Landsat ETM+ TIR data were fused by Spatio-temporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) in order to derive high resolution synthetic data. The synthetic images were

evaluated for both testing and simulated satellite images. The average difference (AD) and absolute average difference (AAD) are smaller than 1.7 K, where the correlation coefficient (CC) and root-mean-square error (RMSE) are 0.755 and 1.824, respectively, showing that the proposed method enhances the spatial resolution of the predicted LST images and preserves the spectral information at the same time.

Keywords: extreme learning machine; Landsat; land surface temperature; MODIS; spatial-temporal fusion; thermal infrared images

1. Introduction

In order to monitor the rapid and continual changes of the global environment, Land Surface Temperature (LST), as the prime and basic physical parameter of the earth's surface, has been studied for over a decade. LST plays a key role in modeling the surface energy balance [1,2] and has a significant impact on analyzing the heat-related issues such as soil moisture [3,4], evapotranspiration [5–7], and urban heat islands [8,9]. Compared with traditional methods using data from weather stations, remote sensing satellite images provide a more effective and efficient method to estimate LST and offer a synoptic view of the study area. However, due to the limitations on both spatial and temporal resolution of the existing satellite sensors, there is no earth observation which can obtain Thermal Infra-Red images (TIR) at detailed spatial- and temporal-resolution simultaneously.

Currently, the applications of thermal infrared remote sensing in urban environment studies require heat-related information at high spatial resolution, as well as high temporal resolution [7]. However, in order to collect more reflected and emitted signal from the earth, large spatial coverage, e.g., lower spatial resolution from the earth observation is required. For instance, several sensors such as Landsat Thematic Mapper (TM)/Enhanced TM Plus (ETM+)/Operational Land Imager (OLI) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), obtain TIR data between 60 m and 120 m [10–13] and they are always used for regional and global LST studies [14,15]. In addition, the platforms with a minimum 16-day revisit cycle may prohibit its application. The revisit cycle for a particular area may be extended due to the poor atmospheric conditions, such as cloud and haze [16]. Particularly in cloud-prone environments, e.g., Hong Kong and the Pearl River Delta region, the poor atmospheric conditions result in a very low probability of obtain cloud-free Landsat imagery (e.g., 10% in a year with cloud cover below 10%) [17].

On the contrary, other sensors, such as Advanced Very High Resolution Radiometer (AVHRR), and MODerate resolution Imaging Spectroradiometer (MODIS), provide a daily revisit cycle, but at a coarser spatial resolution ranging from 250–1000 m, which may not be able to detect the high level of detailed information and seriously impede their potential applications [18,19]. As a consequence, it is necessary to develop a new image fusion method that can integrate complementary characteristics from multi-sensors, in order to generate synthetic LST data with high spatial and temporal resolution.

There are some previous studies of remote sensing data fusion methods using multiple optical sensors [20–22]. Considering different characteristics between the TIR band and the visible bands, traditional image fusion models, such as the widely used Principle Component Analysis (PCA) based

fusion methods [23,24], Intensity-Hue-Saturation (HIS) transformation method [25,26] and wavelet-based image fusion methods [27,28], are more concentrating on the visual effects of the fused images, which may not be useful for quantitative remote sensing applications. Hence, some downscaling approaches have been developed for TIR images. A generalized theoretical framework with semi-empirical regression and modulation integration techniques was constructed by Zhan *et al.* [29] and they evaluated all three fusion data including digital number, radiance and land surface temperature. Rodriguez-Galiano *et al.* [30] applied the Downscaling CoKriging (DCK) method to obtain high spatial resolution LST images using the same scene data at a coarser resolution and the Normalized Difference Vegetation Index (NDVI). The downscaled TIR band was also used to improve the land cover classification accuracy and derive evapotranspiration images of surface energy balance model in a large heterogeneous landscape [31–33].

However, this study is not only aiming on downscaling the TIR images to retrieve higher details of spatial information, but also on fusing high spatial and high temporal data from multi-sensors, e.g., blending high spatial resolution Landsat data and high temporal MODIS data, in order to generate synthetic LST maps. A spatial and temporal data fusion model named the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) [34] was firstly developed to predict daily surface reflectance at a Landsat 30 m spatial resolution using one or more pairs of Landsat and MODIS images on the same day, and one MODIS image on the predicted day, based on the weighted average approach. The STARFM algorithm has been widely used to provide information for monitoring the seasonal changes in vegetation cover and large changes in land use [34]. The performance of STARFM is highly dependent on the characteristic of the landscape patch size. In order to overcome the shortcomings of STARFM, Zhu *et al.* [35] proposed an Enhance STARFM model to improve the accuracy of the predicted image. Compared to STARFM, ESTARFM results preserve more spatial details on finer resolution images, especially for heterogeneous land covers. Hilker *et al.* [36] also designed a new fusion approach named Spatial and Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) for distinguishing and recording the changes of reflectances in both MODIS and Landsat data. Among the above approaches, only STAARH considers the change in land cover types, but it can only be used for predicting the changes in forest [36]. A semi-physical fusion approach employs the MODIS BRDF/Albedo land surface characterization product and Landsat ETM+ data to predict ETM+ reflectance, and the method can also be used for ETM+ cloud/cloud shadow, SLC-off gap filling, and relative radiometric normalization [37]. Through an unmixing-based spatial and temporal fusion model, Zurita Milla *et al.* [38] integrated Landsat TM and MERIS data, however, the results were highly dependent on the quality of the land cover map generated from Landsat images. In addition, a robust SParse representation-based SpatioTemporal reflectance Fusion Model (SPSTFM) has been developed to predict high spatial resolution surface reflectance through data blending with low spatial resolution data. Results show that the SPSTFM is able to capture the changes of surface reflectances in both the changes of phenology and land-cover types [39,40]. These methods were designed for data fusion of surface reflectance rather than for TIR data. Huang *et al.* [41] developed a spatiotemporal image fusion model based on bilateral filtering to derive high resolution LST images. This model works in a densely time-series LST data for urban heat island monitoring. Weng *et al.* [42] proposed a new data fusion model named Spatio-Temporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT) to predict synthetic thermal radiance and LST images at both high temporal and spatial resolution by blending daily MODIS and periodic Landsat datasets. The

prediction accuracy of SADFAT, as measured by the mean absolute difference, ranged from 1.25 K to 2 K. SADFAT was developed for predicting Landsat-like thermal radiance and LST data based on the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) which also considers annual temperature cycle (ATC) and urban thermal landscape heterogeneity [42].

In this paper, in order to generate synthetic high spatial resolution LST images from available satellite-borne sensors, a thermal sharpening method based on extreme learning machine (ELM) algorithm for neural network regression model was firstly adopted, in order to enhance the 60 m Landsat ETM+ TIR band to 30 m resolution. Then, the MODIS LST and enhanced Landsat ETM+ TIR data were blended through Spatio-temporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) (followed by the algorithm developed by Weng *et al.* [42]) to generate synthetic LST data at 30 m resolution.

2. Methodology

2.1. Extreme Learning Machine (ELM) Algorithm

For pattern recognition and image classification, traditional image fusion methods focus on the enhancement of visual effects, including color and texture enhancement; however, for quantitative remote sensing applications and thermal infrared image analysis, the preservation of spectral brightness values is a higher priority. In this paper, an extreme learning machine (ELM) algorithm [43,44] was selected to obtain the internal physical regression relationship between the 60 m Landsat ETM+ TIR image and corresponding 30 m multispectral bands (including visible, near and shortwave infrared bands).

Extreme Learning Machine (ELM) proposed by Huang *et al.* [45] is designed for single-hidden layer feed-forward neural networks (SLFNs) that can adjust the input weights and determine the output weights analytically. Due to the random determination of the input weights and hidden biases, ELM requires numerous of hidden neurons. In practice, the number of hidden neurons should be larger than the number of the variables in dataset, since the useless neurons from the hidden layer will be pruned automatically.

Figure 1 shows the structure of single hidden layer feed forward neural network based on ELM using the activation function, $\mathrm{g}(x)=\mathrm{sig}(\mathbf{w}_i\cdot\mathbf{x}_i+b_i)$.

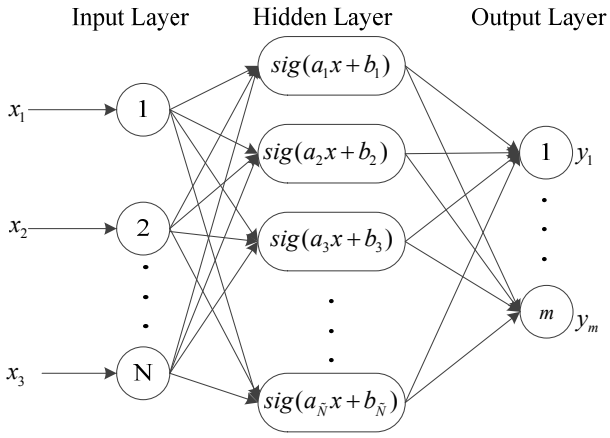

Figure 1. Structure of the single hidden layer feed forward neural network using Extreme Learning Machine (ELM).

The output weight $\boldsymbol{\beta}$ for a training set $A=\left\{\left(\mathbf{x}_i,\mathbf{t}_i\right)\middle|\mathbf{x}_i\in\mathbf{R}^n,\mathbf{t}_i\in\mathbf{R}^m,i=1,2,\cdots,N\right\}$ with activation function $\mathrm{g}(x)$ and hidden neuron $\tilde{N}$ can be calculated according to the following steps:

(i) Allocate the input weight $\mathbf{w}_i$ and bias $b_i,i=1,2,\cdots,\tilde{N}$ arbitrarily.

(ii) Compute the hidden layer output matrix $\mathbf{H}$.

(iii) Use the equation $\boldsymbol{\beta}=\mathbf{H}^{-1}\mathbf{T}$ to calculate the output weight matrix $\boldsymbol{\beta}$.

2.2. Spatio-Temporal Adaptive Data Fusion Algorithm for Temperature Mapping (*SADFAT*)

The Spatio-Temporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT), developed by Weng *et al.* [42], was used to predict synthetic Landsat-like thermal radiance and LST images in this paper. In the algorithm, the MODIS radiance images were first resampled to the same spatial resolution (e.g., 30 m) of Landsat ETM+ images, and *M* is denoted as the MODIS pixel while *L* is denoted as the Landsat pixel. In urban areas, most of the pixels from MODIS images covered more than one land cover type, and these were named mixed pixels. Therefore, the linear spectral mixture analysis (LSMA) algorithm was employed to define the radiance of a mixed pixel. In order to correlate the Landsat radiance with the MODIS radiance, it is assumed that each *L* pixel is deemed as pure pixel and can be considered as an end-member of a *M* pixel; as a result, the radiance of the *M* pixel can be defined as:

$$R_M(t_j)=\sum_{i=1}^{N}l_iR_M\left(x,y,t_j\right)=\sum_{i=1}^{N}l_i\left(\frac{1}{k_1}R_L(x,y,t_j)-\frac{k_2}{k_1}\right) \tag{1}$$

where *R* denotes the radiance at the satellite sensor, *N* is the pixel number of Landsat within a MODIS pixel and l_i denotes the fraction of each Landsat pixel, (*x*, *y*) represents a given location, *t* is the acquisition date, and k_1, k_2 are the coefficients for the relative adjustment for the Landsat and MODIS radiance pixels [42]. Therefore, if there are two pairs of Landsat ETM+ and MODIS image acquired at t_1 and t_2, respectively, Equation (2) indicates that the ratio of the radiance change of j_{th} *L* pixel to the radiance of corresponding *M* pixel is constant for a certain *L* pixel can be quantified as below:

$$\frac{R_{jL}(t_2)-R_{jL}(t_1)}{R_M(t_2)-R_M(t_1)}=\frac{\cos\left(\omega_k+a\bar{e}\right)}{\frac{1}{k_1}\sum_{i=1}^{N}l_i\cos\left(\omega_i+a\bar{e}\right)}=h_j \tag{2}$$

where h_j is denoted as the conversion coefficient for the purpose of consistency, ω reflects the phase shift of a pixel and is related with the thermal characteristics of land cover, and it is a constant if the land surface materials does not change in the period of observation [42]. Thus, if there is a pair of *L* and *M* radiance images at t_0 and an *M* radiance image at t_p, the *L* radiance image at t_p can be predicted by the following equation:

$$R_L(x,y,t_p)=R_L(x,y,t_0)+h(x,y)\times\left[R_M(x,y,t_p)-R_M(x,y,t_0)\right] \tag{3}$$

By adding information from neighbouring spectral similar pixels, a moving window would be used to calculate the radiance of the central pixel [34]. Therefore, supposing s is the moving window size, the predicted *L* pixel radiance can be rewritten as:

$$R_L(x_{s/2}, y_{s/2}, t_p) = R_L(x_{s/2}, y_{s/2}, t_0) + \sum_{i=1}^{N} W_i \times h_i \times \left[R_M\left(x_i, y_i, t_p\right) - R_M\left(x_i, y_i, t_0\right) \right] \quad (4)$$

where W_i represents the weight of a neighboring similar pixel, and N is the number of the spectral similar pixel. In this paper, after predicting the radiance image using SADFAT method, enhanced Landsat ETM+ TIR data with both high spatial and temporal resolution would be converted to LST using the generalized single channel method [46]. More details of SADFAT can be referred to Weng *et al.* [42].

2.3. Implementation of the Proposed Data Fusion Model

Figure 2 presents a flowchart of the proposed fusion model. The implementation consists of two steps to produce synthetic spatial resolution LST maps using Landsat ETM+ and MODIS LST images. In the first step, a thermal sharpening method using ELM algorithm was selected to improve the Landsat ETM+ TIR 60 m resolution image to 30 m resolution. The multispectral and TIR digital number (DN) values were first converted to the radiance values using metadata, and these could ensure the two satellite images have a strong intrinsic correlation of the same land surface types. Then, the hidden layer output matrix, *i.e.*, internal physical regression relationship between the 60 m spatial resolution of Landsat ETM+ TIR band and corresponding six 30 m resolution multispectral bands, was ascertained through the ELM algorithm where the hidden neurons were set to 1000. Finally, the enhanced 30 m resolution Landsat TIR data were computed using the weight vector obtained by neural network regression model.

In the second step, the MODIS LST and enhanced Landsat ETM+ TIR data were blended using SADFAT algorithm [42] to generate a final synthetic LST data at 30 m resolution. The algorithm requires at least two pairs of Landsat and MODIS images acquired at the same date and a set of MODIS images on the desired prediction dates. Before implementing the SADFAT algorithm [42], MODIS LST images should be converted to radiances at the Landsat ETM+ effective thermal wavelength and all the images including enhanced Landsat ETM+ TIR and the converted MODIS radiance images should be geo-registered to the same coordinate system and atmospherically calibrated to the surface radiance. The inputs of this step are two pairs of *L* and *M* images at t_1 and t_2 and one *M* image at the prediction date t_p.

The details of data processing are described as below:

(i) Two *L* images were used to search for the spectrally similar pixels using the method described by Gao *et al.* [34] that define a difference threshold between the central pixel and the neighbouring pixels in a moving window.

(ii) The combined weight and conversion coefficient for each similar pixel were computed. Here, a similar pixel with higher thermal similarity and shorter distance to the central pixel would yield a higher weight and the conversion coefficients were decided by the regression analysis of the similar pixels [35].

(iii) Equation (5) was employed to compute the desired predicted image at t_p. Considering the temporal weights of the two images given by the temporal changes in coarser radiance images, an accurate radiance image can be computed by using the weighted combination of the two predicted radiance images as follows:

$$L(x_{s/2}, y_{s/2}, t_p) = T_{t_1} \times P_{t_1}(x_{s/2}, y_{s/2}, t_p) + T_{t_2} \times P_{t_2}(x_{s/2}, y_{s/2}, t_p) \quad (5)$$

where $(x_{s/2}, y_{s/2})$ represents a location of the central pixel in a moving window, t is the acquisition date, P_{t_1} and P_{t_2} are the predicted radiance image using L image at t_1 and t_2 as the base image, respectively, in which the temporal weight T_k can be calculated as:

$$T_k = \frac{1\Big/\left(\sum_{i=1}^{s}\sum_{j=1}^{s} M(x_i, y_j, t_k, B) - \sum_{i=1}^{w}\sum_{j=1}^{w} M(x_i, y_j, t_p, B)\right)}{\sum_{k=t_1,t_2} 1\Big/\left(\sum_{i=1}^{s}\sum_{j=1}^{s} M(x_i, y_i, t_k, B) - \sum_{i=1}^{w}\sum_{j=1}^{w} M(x_i, y_j, t_p, B)\right)} \quad (6)$$

where $M(x, y, t, B)$ is the resampled radiance at time t of band B.

(iv) Finally, the LST images can be derived using the generalized single channel method.

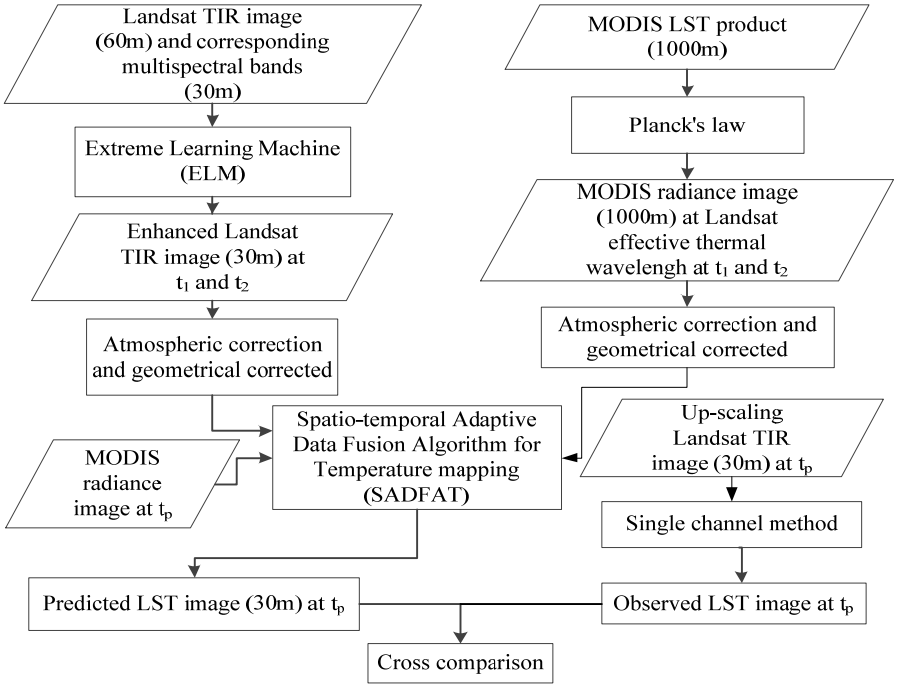

Figure 2. Flowchart of the proposed fusion model for predicting synthetic LST image at 30 m resolution.

3. Results

3.1. Study Area

A study area of 12 km × 12 km, part of Guangzhou, China, was selected (see Figure 3). The study area consists of various landscapes, including water, impervious surface, bare soil, vegetation, *etc.*, and possesses a subtropical climate with moist summers and dry winters.

Three Landsat ETM+ images covering the study area (path/row: 122/44) were obtained from the United States Geological Survey (USGS) website. These datasets were acquired on 20 October, 7 December and 23 December 2013, and they are the L1G level product and are geographically corrected.

The corresponding daily MODIS LST (MOD11A1), reflectance data (MOD09GA) and water vapour content data (MOD05_L2) were downloaded through the Atmosphere Archive and Distribution System website. The MODIS LST (MOD11A1) products were derived from the generalized split-window LST algorithm and the uncertainty of the MODIS-LST production algorithm is around 1 °C in the range of −10 °C to 50 °C, for the surfaces with known emissivities [47].

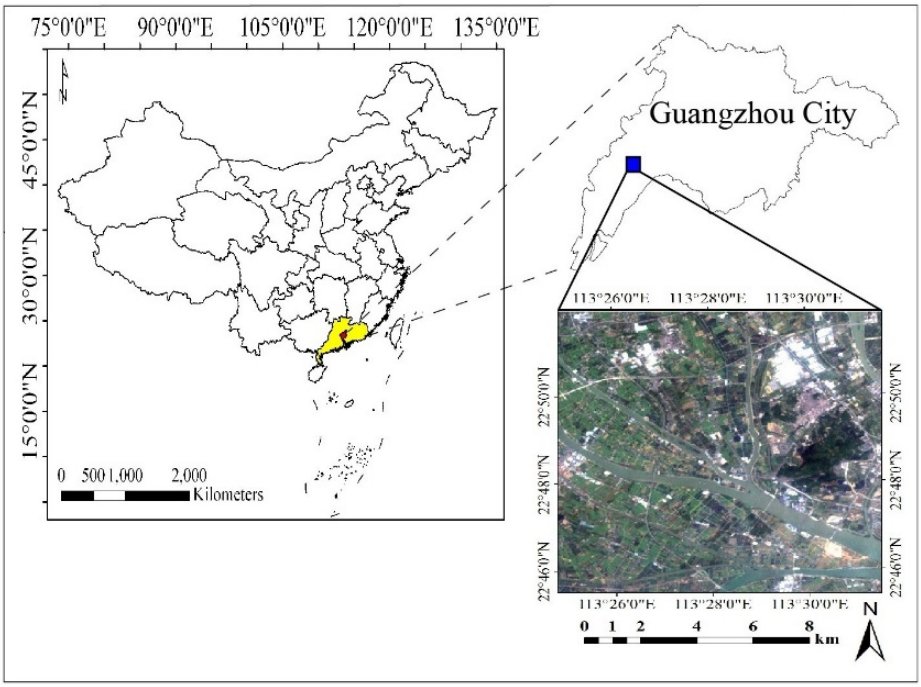

Figure 3. Location of the study area.

3.2. Experiment Results

In this study, the developed fusion methodology was applied to three pairs of Landsat ETM+ and MODIS images both in testing and simulated experiments (Figure 4). Two pairs of the MODIS and Landsat images, acquired on 20 October and 23 December 2013, respectively, were used as base images input to the fusion model, while the remaining images acquired on 7 December 2013, was used to validate the accuracy of prediction. For the testing experiment, the original multispectral images and TIR image of Landsat ETM+ were downscaled to 120 m using the pixel averaging method, then the ELM algorithm was employed to enhance the spatial resolution of the degraded TIR image from 120 m to 60 m, and the original TIR image at 60 m was used as the referenced data for validation. The correlation coefficient (CC), root-mean-square error (RMSE), average difference (AD) and absolute average difference (AAD) were selected as indicators to evaluate the accuracy of thermal spatial sharpening of Landsat TIR images using ELM algorithm and to validate the synthetic LST. For the simulated experiments, since there is no reference to TIR images at 30 m resolution, the reliability of ELM and SADFAT was validated only through visual effects.

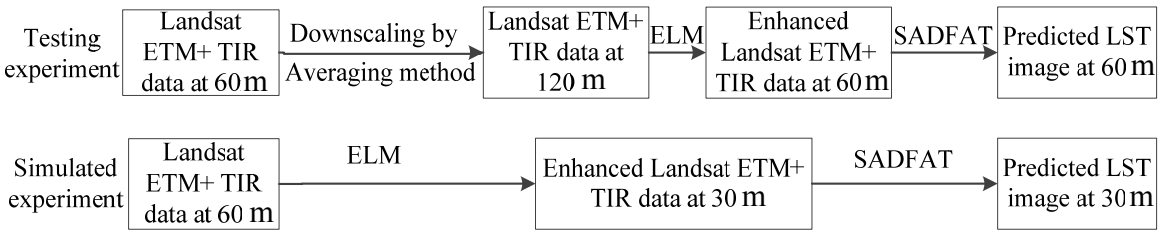

Figure 4. The flowchart of the testing and simulated experiment.

Figures 5 and 6 showed the thermal spatial sharpening results (lower row) of all the testing and simulated Landsat radiance images (middle row) using the extreme learning machine (ELM) algorithm for the neural network regression model. Comparing the spatial sharpened images in both testing and simulated experiments, it can be observed that the sharpened images have more spatial details, such as roads, textural information of forest, urban streets and edges of rivers. Additionally, CC, RMSE, AD, AAD were computed between the sharpened and original radiance images at both 60 m resolution (Table 1). The RMSEs between the sharpened and original radiance images are 0.0844, 0.0891, and 0.0909, and the CC values are 0.8788, 0.8251, and 0.8017 on different dates, respectively. AD values on 20 October and 7 December 2013 are negative values showing that the predictions overestimate the TIR data slightly. Although some details may be lost during the ELM spatial enhancement, the low RMSEs and the strong CCs indicate that the ELM can preserve much of the physical TIR information in the original input TIR images for quantitative remote sensing applications.

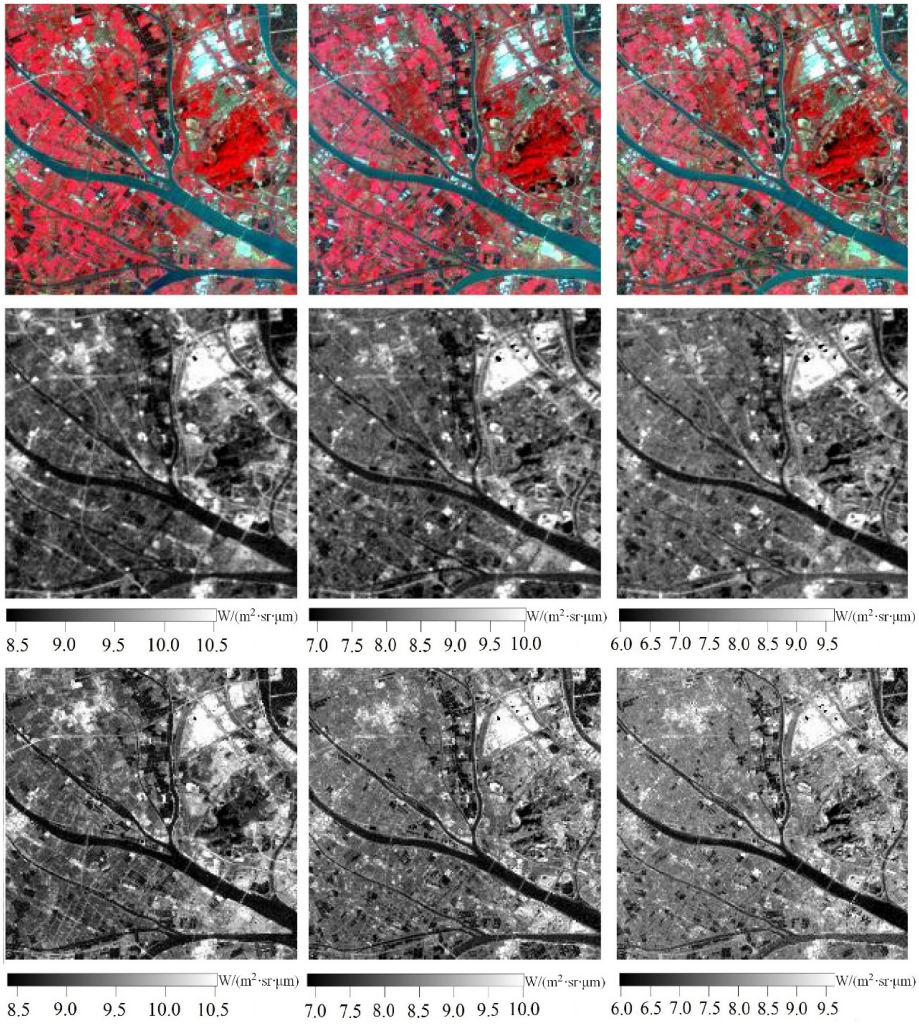

Figure 5. Thermal spatial sharpening results of testing experiments via ELM algorithm. False colour images of downscaled Landsat ETM+ multispectral data at 60 m (**upper row**), downscaled TIR images at 120 m (**middle row**) and sharpened TIR images at 60 m (**lower row**). The original TIR images at 60 m refer to the middle row in Figure 6. From left to right, they were acquired on 20 October, 7 December and 23 December 2013.

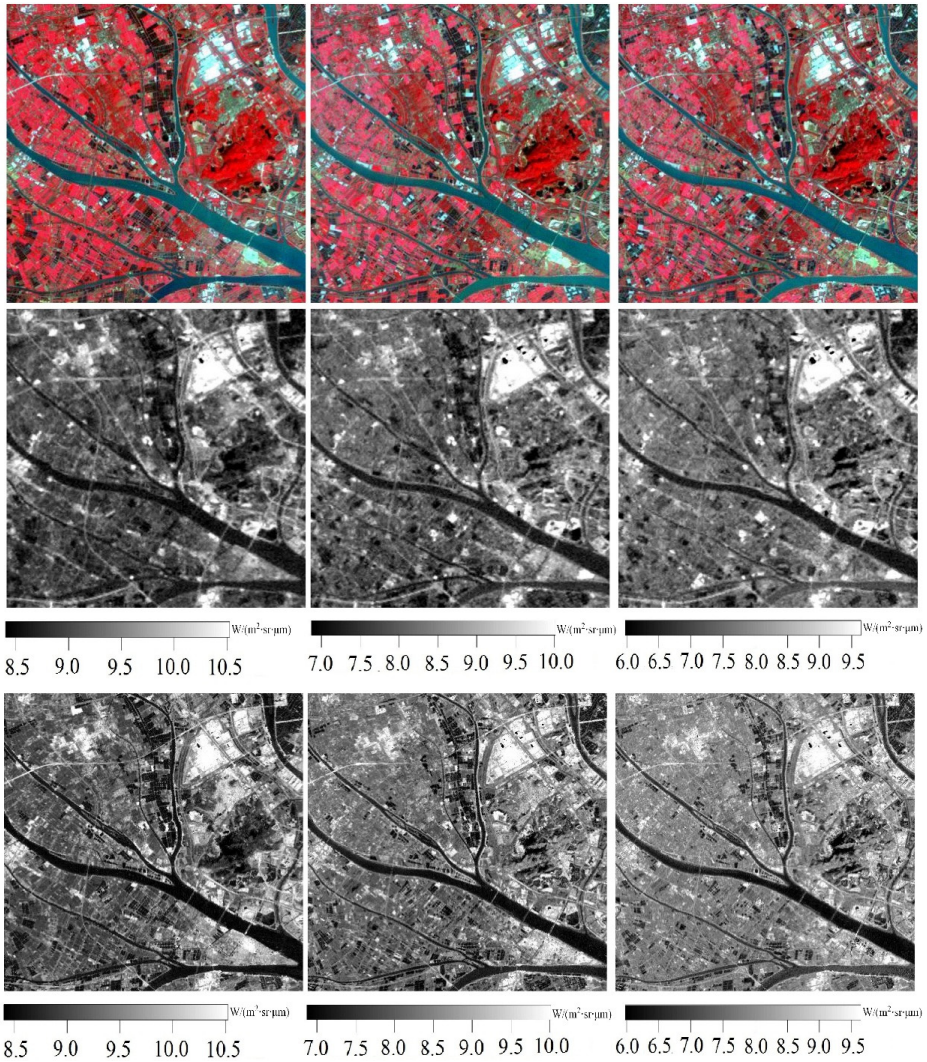

Figure 6. Thermal spatial sharpening results of simulated experiments using ELM algorithm. False colour images of observed Landsat ETM+ multispectral data at 30 m (**upper row**), TIR images at 60 m (**middle row**) and Enhanced TIR images at 30 m (**lower row**). From left to right, they were acquired on 20 October, 7 December and 23 December 2013.

Table 1. Quantitative assessment of thermal sharpening in the testing experiment via ELM.

Date	20 October 2013	7 December 2013	23 December 2013
CC	0.8788 *	0.8251 *	0.8017 *
RMSE	0.0844	0.0891	0.0909
AD	−9.8940e-06	−2.1980e-06	6.3913e-06
AAD	0.0595	0.0620	0.0614

Note: AD is the average of the original value minus the ELM predicted value; * significant at 0.001 level ($p < 0.001$).

In the second step of the proposed fusion model, considering the characteristics of complex surface in this study area, the size of the searching window was set to three MODIS pixels and there are five land cover types in the SADFAT. To ensure the accuracy of searching similar pixels and obtain the NDVI values at the prediction date for LST retrieval, the reflectance values of both Band 3 and Band 4 combined with the thermal radiance information from Band 6 of Landsat ETM+ and its corresponding

bands of MODIS at t_1 and t_2 were input into the SADFAT fusion model developed by Weng *et al.* [42]. Figures 7 and 8 showed the results of predicted images at 60 m and 30 m on 7 December 2013, respectively. For both the testing and simulated experiments, the predicted LST images (Figure 7b and Figure 8b) are visually similar to the original Landsat LST images (Figure 7c and Figure 8c), but they provide higher spatial details in terms of thermal information.

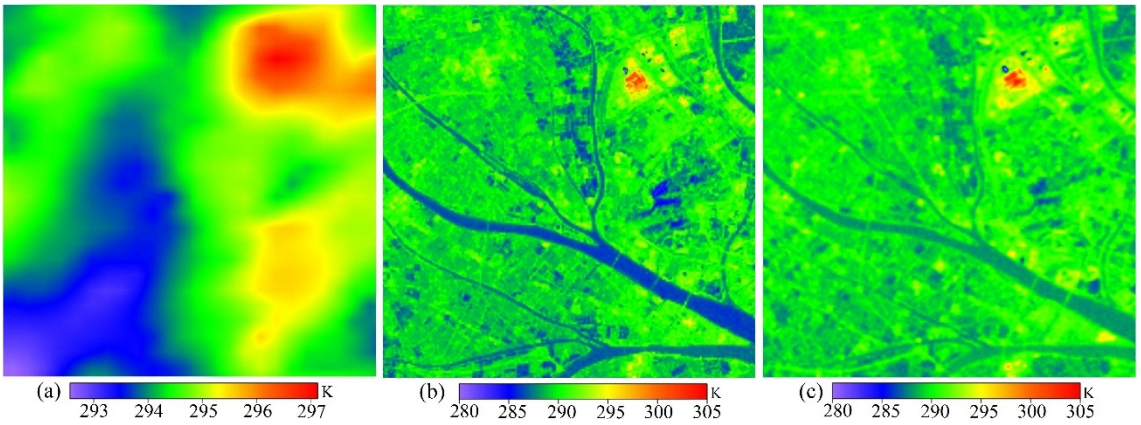

Figure 7. SADFAT result in the testing experiment. **(a)** MODIS LST images resampled from 1 km to 60 m; **(b)** SADFAT-derived image at 60 m; **(c)** original Landsat ETM+ LST image at 60 m spatial on 7 December 2013.

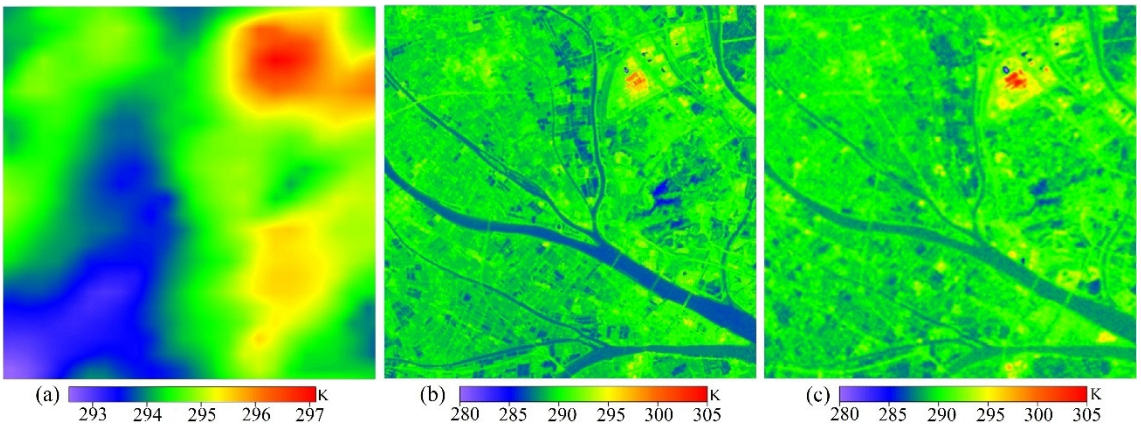

Figure 8. SADFAT result in simulated experiment. **(a)** MODIS LST images resampled from 1 km to 30 m; **(b)** SADFAT-derived image at 30 m; **(c)** up-scaled Landsat ETM+ LST image at 30 m spatial on 7 December 2013.

Figure 9 displays the scatter plots between the predicted and observed LSTs for testing experiment on 7 December 2013. It shows a strong agreement between predicted and observed LSTs in testing experiment. Nevertheless, there are slight differences in some pixels between the predicted and observed LSTs, owing to the limitation of SADFAT that may not be suitable for the changes in land cover and other surface conditions during the prediction period [42]. Table 2 shows the results of CC, RMSE, AD and AAD between the predicted and observed LSTs for the testing experiment at 60 m. The values of AD and AAD are smaller than 2.0 K whereas the CC values are larger than 0.75 suggesting significant consistency.

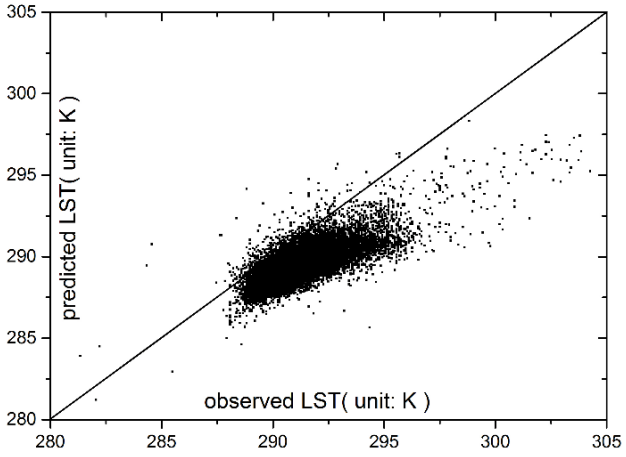

Figure 9. Scatter plots between the predicted and original LSTs at 60 m of the testing experiment on 7 December 2013. *x*-axis denotes the original LSTs, and *y*-axis denotes the prediction LSTs.

Table 2. Indicators of prediction accuracy on 7 December 2013 (unit: K).

Indicator	CC	RMSE	AD	AAD
testing experiment	0.7554 *	1.8242	1.6354	1.6498

Note: AD is the average of the original LST value minus the predicted LST value; *significant at 0.001 level ($p < 0.001$).

4. Discussion

The proposed fusion model employs ELM to enhance the spatial resolution of TIR data before using the spatial-temporal fusion algorithm. As shown in Figures 5 and 6, the sharpened images of both testing and simulated experiments have more spatial details compared with original Landsat TIR images. As shown in Figure 7b and Figure 8b, it can be observed that the predicted LST images provide higher spatial details in terms of thermal information than the original Landsat LST images (Figure 7c and Figure 8c) in both the testing and simulated experiments. In Figure 9, the scatter plot of the testing experiment shows a strong relationship between predicted and original LSTs at 60 m. In addition, small values of AD and AAD (Table 2) (less than 1.7 K) and high values of CC (larger than 0.75) are observed. These results show that our proposed method can enhance the spatial resolution of the predicted LST image and also preserve the spectral information simultaneously.

However, there are still some limitations/errors in the proposed method, including: (i) the thermal spatial sharpening ELM algorithm is highly time-consuming; (ii) some important information in the original TIR images may be lost during the ELM spatial enhancement; (iii) the proposed fusion method cannot predict the changes in LST that are not presented on the MODIS and/or Landsat images, especially in the missing pixels caused by cloud contamination [42]; (iv) the proposed fusion method requires at least two input pairs of fine and coarse resolution images, in the same season or under similar atmospheric conditions; (v) the number of land cover types and the size of a moving window should be determined by searching similar pixels in SADFAT, which may limit the automated process. Thus, more research on the following issues will be conducted in the near future: (i) by improving the fusion accuracy and efficiency of ELM simultaneously, and (ii) developing a more advanced fusion method for generating synthetic LST images to resolve the limitations of seasonal changes and different atmospheric conditions.

5. Conclusion

In this paper, a novel data fusion method by integrating a thermal spatial sharpening algorithm with spatial-temporal fusion to generate synthetic LST datasets at both high spatial and temporal revolution was developed. The performance of this method was tested using Landsat ETM+ and MODIS images both in testing and simulated experiments. At first, the extreme learning machine algorithm was selected to enhance the spatial resolution of the Landsat ETM+ TIR data. Then, the MODIS LST and enhanced Landsat ETM+ TIR data were fused using SADFAT developed by Weng *et al.* [42] to derive high temporal resolution synthetic data. The proposed image fusion method provides an alternative to generate the synthetic high resolution image for remote sensing applications from multi-source satellite data. Compared with the traditional spatio-temporal adaptive data fusion algorithm, the synthetic LST images derived in this study can depict more spatial details. The generated synthetic LST product can be used for monitoring the variation of land surface temperature in urban heat island studies. Further research on other sensors such as Landsat TIRS, HJ-1B satellite will be conducted.

Acknowledgments

This research was sponsored by the grant PolyU 1-ZVBR from the Research Institute for Sustainable Urban Development, the Hong Kong Polytechnic University, grant G-YM85 from the Hong Kong Polytechnic University, and grant F-PP1Q from the Early Career Scheme of the Hong Kong Research Grants Council. This work was also jointly supported by the National Basic Research Program (973 Program, Grant No.2011CB707102), and the project funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions. The authors wish to thank Prof. Qihao Weng, Dr. Feng Gao, and Dr. Xiaolin Zhu for their valuable suggestions in this study; We would also like to thank the NASA Goddard Space Flight Centre and U.S. Geological Survey for the MODIS and Landsat images.

Author Contributions

Yai Bai, Man Sing Wong and Kai Qin conceived and designed the experiments, performed the experiments and analyzed the data; Yai Bai, Man Sing Wong, Wen-zhong Shi, Li-xin Wu and Kai Qin prepared the manuscript.

Appendix A. Extreme Learning Machine (ELM) Algorithm

Extreme Learning Machine (ELM) proposed by Huang *et al.* [45] is designed for single-hidden layer feed-forward neural networks (SLFNs) that can adjust the input weights and determine the output weights analytically. Supposing the number of training samples $(\mathbf{x_i}, \mathbf{t_i})$ is N, where $\mathbf{x}_\mathbf{i} = [x_{i1}, x_{i2}, \cdots, x_{in}]^{\mathrm{T}} \in \mathbf{R}^n$ and $\mathbf{t}_\mathbf{i} = [t_{i1}, t_{i2}, \cdots, t_{im}]^{\mathrm{T}} \in \mathbf{R}^m$, the standard SLFNs equipped with $\tilde{N}$ hidden neurons and activation function $\mathrm{g}(x) = \mathrm{sig}(\mathbf{w}_i \cdot \mathbf{x}_i + b_i)$ can be mathematically expressed as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{7}$$

where $\mathbf{H}\left(\mathbf{w}_1,\cdots,\mathbf{w}_{\tilde{N}},b_1,\cdots,b_{\tilde{N}},\mathbf{x}_1,\cdots,\mathbf{x}_{\tilde{N}}\right)=\begin{bmatrix} \mathrm{g}\left(\mathbf{w}_1\cdot\mathbf{x}_1+b_1\right) & \cdots & \mathrm{g}\left(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_1+b_{\tilde{N}}\right) \\ \vdots & \cdots & \vdots \\ \mathrm{g}\left(\mathbf{w}_1\cdot\mathbf{x}_N+b_1\right) & \cdots & \mathrm{g}\left(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_N+b_{\tilde{N}}\right) \end{bmatrix}_{N\times\tilde{N}}$, $\boldsymbol{\beta}=\begin{bmatrix} \boldsymbol{\beta}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\beta}_{\tilde{N}}^{\mathrm{T}} \end{bmatrix}_{\tilde{N}\times m}$ and $\mathbf{T}=\begin{bmatrix} \mathbf{t}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{t}_N^{\mathrm{T}} \end{bmatrix}_{N\times m}$, And $\mathbf{H}$ represents the hidden layer output matrix of the neural network; b_i is the threshold of the i_{th} hidden neuron; $\mathbf{w}_i=\left[w_{i1},w_{i2},\cdots,w_{in}\right]^{\mathrm{T}}$ denotes the weight vector connecting the i_{th} hidden neuron and the input neurons; $\boldsymbol{\beta}_i=\left[\beta_{i1},\beta_{i2},\mathrm{L},\beta_{im}\right]^{\mathrm{T}}$ stands for the weight vector connecting the i_{th} hidden neuron and the output neurons.

Appendix B. Spatio-Temporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT)

The Spatio-Temporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT), developed by Weng *et al.* [42], was used in this study to predict synthetic thermal radiance and LST images in the second step. Below is a brief description of the algorithm:

If there are two pairs of Landsat ETM+ and MODIS image acquired at t_1 and t_2, respectively, the changes of radiance of a M pixel between t_1 and t_2 can be computed as:

$$R_M(t_2)-R_M(t_1)=\frac{1}{k_1}\sum_{i=1}^{N}l_i\left(R_{iL}(t_2)-R_{iL}(t_1)\right) \tag{8}$$

where R denotes the radiance at the satellite sensor, N is the pixel number of Landsat within a MODIS pixel and l_i denotes the fraction of each Landsat pixel, t is the acquisition date, and k_1 is the coefficient for the relative adjustment for the Landsat and MODIS radiance pixels. Considering the seasonal change of LST based on ATC, and through the Planck's law, the radiance change of an L pixel from t_1 to t_2 can be expressed as:

$$R_{iL}(t_2)-R_{iL}(t_1)=2d\cos\left(\omega_i+a\frac{e_1+e_2}{2}\right)\sin\left(a\frac{e_2-e_1}{2}\right)=D\cos\left(\omega_i+a\bar{e}\right) \tag{9}$$

where a is the angular frequency, ω is the phase shift, or heat lag, d is the amplitude of the radiance variation, D is the constant, and $\bar{e}$ is the mean acquisition date, e_1 and e_2 are the parameters input to the algorithm. Therefore, if the radiances of the j_{th} L pixel at date t_1 and t_2 are known, the equation can be formulated as Equation (10):

$$R_{jL}(t_2)-R_{jL}(t_1)=2d\cos\left(\omega_i+a\bar{e}\right)\sin\left(a\frac{e_2-e_1}{2}\right) \tag{10}$$

By combining Equation (9) with Equation (10), the Equation (11) indicates that the ratio of the radiance change of j_{th} L pixel to the radiance of the corresponding M pixel is constant for a certain L pixel, which can be quantified as below:

$$\frac{R_{jL}(t_2)-R_{jL}(t_1)}{R_M(t_2)-R_M(t_1)}=\frac{\cos\left(\omega_k+a\bar{e}\right)}{\frac{1}{k_1}\sum_{i=1}^{N}l_i\cos\left(\omega_i+a\bar{e}\right)}=h_j \tag{11}$$

where h_j is named the conversion coefficient for the purpose of consistency, ω reflects the phase shift of a pixel and is related with the thermal characteristics of land cover.

Supposing s is the moving window size, if there is a pair of *L* and *M* radiance images at t_0 and a *M* radiance image at t_p, the predicted *L* radiance image at t_p can be predicted by the following equation:

$$R_L(x_{s/2},y_{s/2},t_p)=R_L(x_{s/2},y_{s/2},t_0)+\sum_{i=1}^{N}W_i\times h_i\times\left[R_M\left(x_i,y_i,t_p\right)-R_M\left(x_i,y_i,t_0\right)\right] \tag{12}$$

where W_i represents the weight of a neighbouring similar pixel, and N is the number of the spectral similar pixel. In this paper, after predicting the radiance image using SADFAT method (Weng *et al.* [42]), enhanced Landsat ETM+ TIR data with both high spatial and temporal resolution would be converted to LST using the generalized single channel method [46]. The following equations are the implementation procedures of the generalized single channel method for ETM+ TIR data.

$$T_{surface}=\alpha\left[\varepsilon^{-1}\left(\varphi_1R_{sensor}+\varphi_2\right)+\varphi_3\right]+\beta \tag{13}$$

$$\alpha=\left(\frac{c_2R_{sensor}}{T_{sensor}^2}\left(\frac{\lambda^4}{c_1}R_{sensor}+\lambda^{-1}\right)\right)^{-1}\text{ and }\beta=-\alpha R_{sensor}+T_{sensor} \tag{14}$$

where $T_{surface}$ is the retrieval temperature of the surface, R_{sensor} and T_{sensor} stands for the radiance and brightness temperature at-satellite, respectively, λ (11.3355 μm) is the effective wavelength for ETM+ sensor, c_1, c_2 are the constants, ε is the land surface emissivity. The atmospheric functions are described with the water vapour content:

$$\begin{bmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\end{bmatrix}=\begin{bmatrix}0.14714 & -0.15583 & 1.1234\\ -1.1836 & -0.37607 & -0.52894\\ -0.04554 & 1.8719 & -0.39071\end{bmatrix}\begin{bmatrix}\upsilon^2\\ \upsilon\\ 1\end{bmatrix} \tag{15}$$

where υ is the water vapour content, which can be obtained from the water vapour content product (MOD05_L2).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kalma, J.D.; McVicar, T.R.; McCabe, M.F. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys Geophys.* **2008**, *29*, 421–469.
2. Cammalleri, C.; Anderson, M.; Ciraolo, G.; D'Urso, G.; Kustas, W.; La Loggia, G.; Minacapilli, M. Applications of a remote sensing-based two-source energy balance algorithm for mapping surface fluxes without *in situ* air temperature observations. *Remote Sens. Environ.* **2012**, *124*, 502–515.

3. Srivastava, P.K.; Han, D.; Ramirez, M.R.; Islam, T. Machine learning techniques for downscaling smos satellite soil moisture using modis land surface temperature for hydrological application. *Water Resour. Manag.* **2013**, *27*, 3127–3144.
4. Song, X.; Leng, P.; Li, X.; Li, X.; Ma, J. Retrieval of daily evolution of soil moisture from satellite-derived land surface temperature and net surface shortwave radiation. *Int. J. Remote Sens.* **2013**, *34*, 3289–3298.
5. Bateni, S.; Entekhabi, D.; Castelli, F. Mapping evaporation and estimation of surface control of evaporation using remotely sensed land surface temperature from a constellation of satellites. *Water Resour. Res.* **2013**, *49*, 950–968.
6. Tang, R.; Li, Z.-L.; Jia, Y.; Li, C.; Chen, K.-S.; Sun, X.; Lou, J. Evaluating one-and two-source energy balance models in estimating surface evapotranspiration from Landsat-derived surface temperature and field measurements. *Int. J. Remote Sens.* **2013**, *34*, 3299–3313.
7. Anderson, M.; Kustas, W.; Norman, J.; Hain, C.; Mecikalski, J.; Schultz, L.; González-Dugo, M.; Cammalleri, C.; D'Urso, G.; Pimstein, A. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 223–239.
8. Wong, M.S.; Nichol, J.E. Spatial variability of frontal area index and its relationship with urban heat island intensity. *Int. J. Remote Sens.* **2013**, *34*, 885–896.
9. Weng, Q.; Fu, P. Modeling annual parameters of clear-sky land surface temperature variations and evaluating the impact of cloud cover using time series of landsat tir data. *Remote Sens. Environ.* **2014**, *140*, 267–278.
10. Goward, S.N.; Masek, J.G.; Williams, D.L.; Irons, J.R.; Thompson, R. The landsat 7 mission: Terrestrial research and applications for the 21st century. *Remote Sens. Environ.* **2001**, *78*, 3–12.
11. Anderson, M.C.; Allen, R.G.; Morse, A.; Kustas, W.P. Use of landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sens. Environ.* **2012**, *122*, 50–65.
12. Almeida, T.; de Souza Filho, C.; Rossetto, R. Aster and landsat ETM+ images applied to sugarcane yield forecast. *Int. J. Remote Sens.* **2006**, *27*, 4057–4069.
13. Sameen, M.I.; Al Kubaisy, M.A. Automatic surface temperature mapping in arcgis using landsat-8 tirs and envi tools, case study: Al Habbaniyah Lake. *J. Environ. Earth Sci.* **2014**, *4*, 12–17.
14. Li, Y.-Y.; Zhang, H.; Kainz, W. Monitoring patterns of urban heat islands of the fast-growing shanghai metropolis, china: Using time-series of landsat tm/etm+ data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 127–138.
15. Masek, J.G.; Collatz, G.J. Estimating forest carbon fluxes in a disturbed southeastern landscape: Integration of remote sensing, forest inventory, and biogeochemical modeling. *J. Geophys. Res.* **2006**, *111*, doi:10.1029/2005JG000062.
16. Ju, J.; Roy, D.P. The availability of cloud-free landsat ETM+ data over the conterminous united states and globally. *Remote Sens. Environ.* **2008**, *112*, 1196–1211.
17. Leckie, D.G. Advances in remote sensing technologies for forest surveys and management. *Can. J. For. Res.* **1990**, *20*, 464–483.
18. Justice, C.O.; Vermote, E.; Townshend, J.R.; Defries, R.; Roy, D.P.; Hall, D.K.; Salomonson, V.V.; Privette, J.L.; Riggs, G.; Strahler, A. The moderate resolution imaging spectroradiometer (modis):

Land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1228–1249.

19. Stathopoulou, M.; Cartalis, C. Downscaling avhrr land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sens. Environ.* **2009**, *113*, 2592–2605.
20. Pohl, C.; Van Genderen, J. Review article multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854.
21. Smith, M.I.; Heather, J.P. A review of image fusion technology in 2005. In *Defense and Security, 2005*; International Society for Optics and Photonics: Bellingham WA, USA, 2005; pp. 29–45.
22. Ha, W.; Gowda, P.H.; Howell, T.A. A review of potential image fusion methods for remote sensing-based irrigation management: Part II. *Irrig. Sci.* **2013**, *31*, 851–869.
23. González-Audícana, M.; Saleta, J.L.; Catalán, R.G.; García, R. Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1291–1299.
24. Naidu, V.; Raol, J. Pixel-level image fusion using wavelets and principal component analysis. *Def. Sci. J.* **2008**, *58*, 338–352.
25. Tu, T.-M.; Huang, P.S.; Hung, C.-L.; Chang, C.-P. A fast intensity-hue-saturation fusion technique with spectral adjustment for ikonos imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 309–312.
26. Choi, M. A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1672–1682.
27. Amolins, K.; Zhang, Y.; Dare, P. Wavelet based image fusion techniques—An introduction, review and comparison. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 249–263.
28. Zhang, Y.; Hong, G. An ihs and wavelet integrated approach to improve pan-sharpening visual quality of natural colour ikonos and quickbird images. *Inf. Fus.***2005**, *6*, 225–234.
29. Zhan, W.; Chen, Y.; Zhou, J.; Li, J.; Liu, W. Sharpening thermal imageries: A generalized theoretical framework from an assimilation perspective. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 773–789.
30. Rodriguez-Galiano, V.; Pardo-Iguzquiza, E.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Downscaling Landsat 7 ETM+ thermal imagery using land surface temperature and NDVI images. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 515–527.
31. Rodriguez-Galiano, V.; Ghimire, B.; Pardo-Igúzquiza, E.; Chica-Olmo, M.; Congalton, R. Incorporating the downscaled landsat tm thermal band in land-cover classification using random forest. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 129–137.
32. Hong, S.-H.; Hendrickx, J.M.; Borchers, B. Up-scaling of sebal derived evapotranspiration maps from Landsat (30m) to MODIS (250m) scale. *J. Hydrol.* **2009**, *370*, 122–138.
33. Jeganathan, C.; Hamm, N.; Mukherjee, S.; Atkinson, P.M.; Raju, P.; Dadhwal, V. Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 178–191.
34. 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.

35. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623.
36. Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of dense time series synthetic landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sens. Environ.* **2009**, *113*, 1988–1999.
37. Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal modis—Landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130.
38. Zurita-Milla, R.; Clevers, J.G.; Schaepman, M.E. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 453–457.
39. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716.
40. Song, H.; Huang, B. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1883–1896.
41. Huang, B.; Wang, J.; Song, H.; Fu, D.; Wong, K. Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1011–1015.
42. Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at landsat resolution by fusing Landsat and MODIS data. *Remote Sens. Environ.* **2014**, *145*, 55–67.
43. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man. Cybern. B Cybern.* **2012**, *42*, 513–529.
44. Yao, W.; Han, M. Fusion of thermal infrared and multispectral remote sensing images via neural network regression. *J. Image Gr.* **2010**, *15*, 1278–1284.
45. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; pp. 985–990.
46. Jiménez-Muñoz, J.C.; Sobrino, J.A. A generalized single-channel method for retrieving land surface temperature from remote sensing data. *J. Geophys. Res.* **2003**, *108*, doi:10.1029/2003JD003480.
47. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.-L. Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote Sens.* **2004**, *25*, 261–274.