Third International Conference on "Materials, Computing and Communication Technologies"
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Image Fusion Using Deep Convolutional Neural Networks

Jeba Jasmine S, Maria Seraphin Sujitha S
Department of Electronics and Communication Engineering, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil, Tamil Nadu, India

Article Info
Volume 8, Issue 7
Page Number : 103-112

Publication Issue :
May-June-2022

Article History
Accepted: 01 June 2022
Published: 20 June 2022

ABSTRACT

Remote sensing images with various resolutions such as MODIS and Landsat images are captured by many earth observing satellites. To get the clear data from the input images, image fusion technology is commonly adopted to generate remote sensing images at high resolution. Here based on deep convolutional neural network, a remote sensing image fusion method that can adequately extract the spatial features from the input source image is proposed. The proposed method consists of the training stage and prediction stage separately. Besides residual learning is adopted between the high and low resolution MODIS images. The proposed method consists of extracting the features by illumination and bicubic interpolation. Second, the feature fusion is done by super resolution convolutional neural network. In the prediction stage, the fused image is compared with other predefined data. Finally, proposed method provides better results compared with other classical methods.

Index Terms—High resolution, Residual learning, Bicubic interpolation

I. INTRODUCTION

Remote sensing simply means obtaining information about an object without touching the object itself. It has two facets; acquiring data by a device at a distance from the object and analyzing data of the object to interpret its physical properties. These two aspects are closely connected to each other. The basic fact in remote sensing is that different wavelengthranges of the electromagnetic spectrum is reflected or emitted from an object at certain intensity, which is dependent upon the physical and compositional attributes of the object.

Remote sensing today plays an important role in geological analysis of large areas which utilizes electromagnetic spectrum not only within the visible range but also beyond the visible range that human eye can't perceive. The unique spectral signatures of rocks, minerals and other geological elements are used to map these geological elements in large areas in a short time using remote sensing data. Earth observation systems generally include infrared region of the electromagnetic spectrum, which include the Visible Near Infrared

(VNIR) and Shortwave Infrared (SWIR). Further some imaging systems such as LANDSAT and ASTER cover Thermal Infrared (TIR) region, which is a wave infrared region in the spectrum.

TIR radiance values of objects can also be used for mapping similar to VNIR and SWIR. As useful as it may be remote sensing like any tool requires continuously increasing improvement. Similarly advances in the technology necessitate the improvement of the methods accordingly, both in terms of accuracy and precision. Image fusion is one of the techniques that are employed to increase spatial and/or spectral resolution of remotely sensed data by fusing a high spatial but low spectral resolution image with a low spatial high spectral resolution image.

The purpose of this study is to fuse TIR and SWIR bands of ASTER with VNIR bands, while evaluating the multispectral infrared data with increased resolution for lithological discrimination and mapping using the basic image fusion techniques. To accomplish these objectives, a Graphical User Interface (GUI) was prepared using the commercial software MATLAB and its image processing toolbox which contains commands and utilities that are commonly used in image processing applications.

Image Fusion (IF) is an emerging field for generating an informative image with the integration of images obtained by different sensors for decision making. The analytical and visual image quality can be improved by integrating different images. Effective image fusion is capable of preserving vital information by extracting all important Information from the images without producing any inconsistencies in the output image. After fusion, the fused image is more suitable for the machine and human perception.

The first step of fusion is Image Registration (IR) in which source image is mapped with respect to the reference image. This type of mapping is performed to match the equivalent image on the basis of confident features for further analysis. IF and IR are perceived as vital assistants to produce valuable information in several domains.

II. RELATED WORK

Andrew A et al.,[1] proposed multiscale analysis of relationship between imperviousness and urban tree height using airborne remote sensing. The relationship between impervious land cover and tree development is of important to understanding urban ecological systems. While impervious surfaces are associated with degraded soil conditions, rerouted hydrological networks and urban microclimates, the overall impact of these effects on tree development is highly variable. This study examines this relationship at two spatial scales: within the individual tree's local environment and across the broad scale urban landscape. Using a fusion of airborne hyperspectral imagery and light detection and ranging (LiDAR) data, a 1.0 m spatial resolution classified land cover map (accuracy of 88.6%) was produced for the city of Surrey, British Columbia, Canada, from which landscape imperviousness was then derived. The stem heights of 1914 trees were estimated from the LiDAR data, to which species specific height models were fit using planting dates recorded by city authorities. Having accounted for the age of the trees, the residuals from these models (i.e.: the difference between modelled and measured height) were then used as indicators of tree development. When aggregated to 0.5 km2 spatial units, negative relationships were found between height model residuals and the degree of land cover imperviousness.

Beena Matikainen et al.,[2] proposed object based analysis of multispectral airborne laser scanner data for land cover classification and map updating. During the last 20 years, Airborne Laser Scanning (ALS), often combined with passive multispectral information from aerial images, has shown its high feasibility for

automated mapping processes. The main benefits have been achieved in the mapping of elevated objects such as buildings and trees. Recently, the first multispectral airborne laser scanners have been launched and active multi spectral information is for the first time available for 3D ALS point clouds from a single sensor. This article discusses the potential of this new technology in map updating, especially in automated object based land cover classification and change detection in a suburban area. For our study, Optech Titan multispectral ALS data over a suburban area in Finland were acquired. Results from an object based random forests analysis suggest that the multispectral ALS data are very useful for land cover classification, considering both elevated classes and ground level classes. The overall accuracy of the land cover classification results with six classes was 96% compared with validation points. The classes under study included building, tree, asphalt, gravel, rocky area and low vegetation. Compared to classification of single channel data, the main improvements were achieved for ground level classes. According to feature importance analyses, multispectral intensity features based on several channels were more useful than those based on one channel.

Hang R et al [3] proposed learning multiscale deep features for high resolution satellite image scene classification, a multiscale deep feature learning method for high resolution satellite image scene classification. Specifically, first warp the original satellite image into multiple different scales. The images in each scale are employed to train a Deep Convolutional Neural Network (DCNN). However, simultaneously training multiple DCNNs is time consuming. To address this issue, this deep feature learning method thus explore DCNN with spatial pyramid pooling (SPP net). Since different SPP nets have the same number of parameters, which share the identical initial values and only fine tuning the parameters in fully connected layers ensures the effectiveness of each network, thereby greatly accelerating the training process. Then, the multiscale satellite images are fed into their corresponding SPP nets respectively, to extract multiscale deep features. Finally, a multiple kernel learning method is developed to automatically learn the optimal combination of such features. Experiments on two difficult data sets show that the proposed method achieves favorable performance compared with other state of the art methods.

Jodriges L et al.,[4] proposed hyper spectral image classification using deep pixel pair features. The deep convolutional neural network (CNN) is of great interest recently. It can provide excellent performance in hyper spectral image classification when the number of training samples is sufficiently large. In this paper, a novel pixel pair method is proposed to significantly increase such a number, ensuring that the advantage of CNN can be actually offered. For a testing pixel, pixel pairs, constructed by combining the center pixel and each of the surrounding pixel, are classified by the trained CNN and the final label is then determined by a voting strategy. The proposed method utilizing deep CNN to learn pixel pair features is expected to have more discriminative power. Experimental results based on several hyperspectral image data sets demonstrate that the proposed method can achieve better classification performance than the conventional deep learning based method.

Magnus G et al.,[5] proposed multispectral and hyper spectral image fusion using a 3D convolutional neural network, a method using a 3D convolutional neural network to fuse together multispectral and hyper spectral images to obtain a high resolution HS image. Dimensionality reduction of the HS image is performed prior to fusion in order to significantly reduce the computational time and make the method more robust to noise. Experiments are performed on a data set simulated using a real HS image. The results obtained show that the proposed approach is very promising when compared with conventional methods. This is especially true when the HS image is corrupted by additive noise.

III. METHODOLOGY

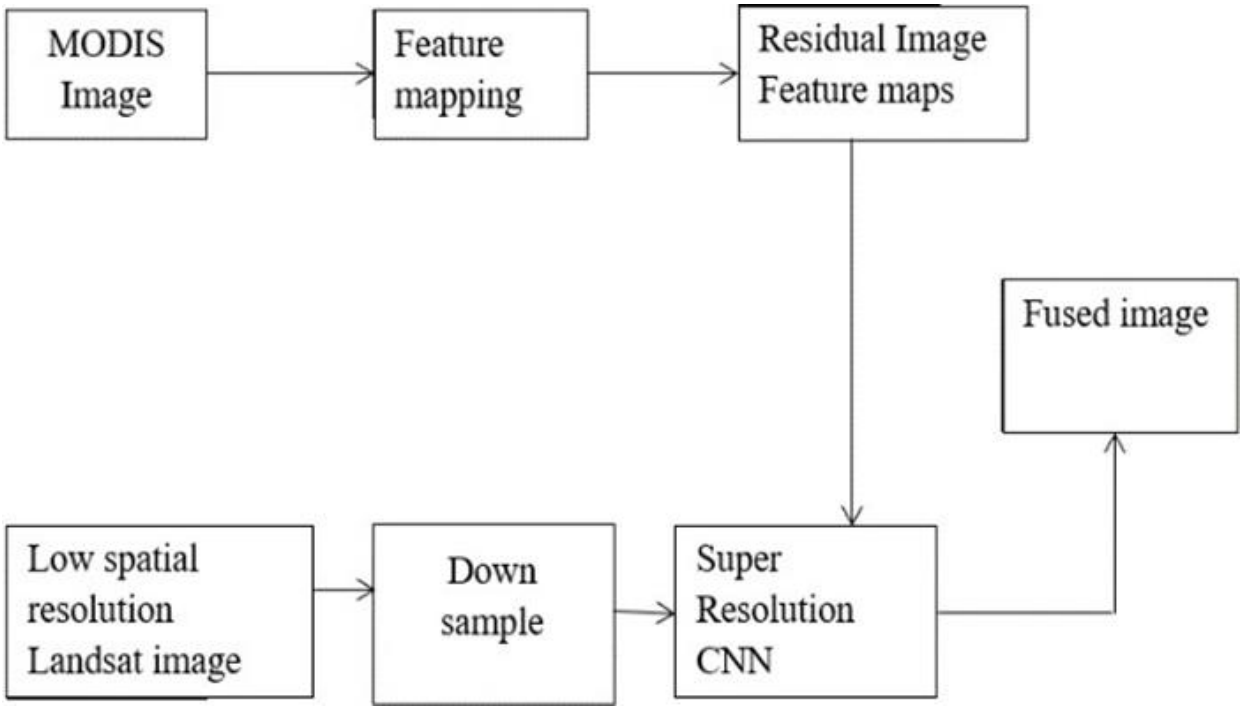


Figure. 1 Block diagram of proposed system

Figure1 shows a fusion model consisting of high pass modulation and weighting is adopted to make full use of the available information. Specifically, the MODIS images in prior and prediction dates are fed into the learned NLM CNN to obtain the transitional LSR Landsat images, which are then fed into the fusion model to get the LSR Landsat image on prediction date. Together with the simulated LSR Landsat images on prior dates, the fusion result in the last step is fed into the learned SR CNN to obtain the transitional Landsat images, which are then fed into the fusion model to get the final fusion result on the prediction date.

A. Training stage

To build a nonlinear model between MODIS and Landsat images, first down samples the spatial resolution of Landsat images to be similar with that of MODIS images. To narrow the resolution gap in the next SR step, the spatial resolution of Landsat images is reduced by two times. Denote the training samples of MODIS and LSR Landsat images as X and Y1, respectively, expect to learn an NLM function.

In the training stage, in order to build a nonlinear mapping model between MODIS and Landsat MODIS residual images, first up sample the spatial resolution of the same size as. Then, the Landsat and MODIS images at the same time are differenced to obtain a residual image. Thus, expect to learn a mapping function which approximates. Pixel value in is likely to be zero or small. To predict this residual image. The loss function now becomes, where the network prediction is made. Divide the high and low resolution images corresponding on the same time into overlapping image patches. Define the set of high and low resolution samples as and where the corresponding samples are land. The overlapping segmentation is performed here to increase the number of training samples. After predicting the residual image, the ground truth Landsat image is obtained by the sum of the input MODIS image and the predicted residual image. In the network, the loss layer has three inputs:

residual estimation, input MODIS image, and Landsat image. The loss is calculated as the Euclidean distance between the reconstructed image and the real Landsat image.

In order to achieve the purpose of high precision spatiotemporal fusion, use a very deep convolutional network. Use 18 layers where layers except the first and the last are of the same type: 64 filters of the size, where a filter operates on spatial region across 64 channels. The first layer operates on the input image. The last layer, used for image reconstruction consists of a single filter of size Considering that X and Y1 are largely similar, the residual image between X and Y1 from X, so focus on learning the high frequency details of Y1. The residual image here of is defined as Rs = Y -Y1. Here Y is the original Landsat image. Thus, learn a mapping function approximates Rs. After predicting the residual image, the ground truth LSR Landsat image is obtained by the sum of the input MODIS image and the residual image.

The recent works have demonstrated the effectiveness of residual learning in image denoising and image SR. As this use demonstrated in the structure of the NLM CNN consists of five layers, the input layer, three convolutional hidden layers, and the output layer, where three hidden layers correspond to three operations: feature extraction NLM and reconstruction. Next, describe these three operations in detail. To extract the features of the input MODIS images, apply n1 filters with kernel size of k1×k1 to each of them. To speed up the convergence of the network while ensuring the accuracy, the rectified linear units [ReLU, max(0)] are adopted for nonlinearity in the filter responses.

Then, n1 MODIS feature maps for each input image are obtained from the first hidden layer. This step can be expressed by the following equation

( ) ( ) (1)

Minimizing the loss between the reconstructed residual images and the corresponding ground truth residual images Rs

= Y – (2)

Given N MODIS and LSR Landsat training samples, the mean squared error (MSE) is adopted as the loss function.

= Σ

(

) - (-)(3)

This loss is minimized by adopting stochastic gradient descent with the standard back propagation. Empirically set the learning rate to for the first two hidden are layers and for the last hidden layer.

For the input MODIS images, first map them to the LSR Landsat images via the NLM CNN and further super resolve the LSR Landsat images to Landsat images via the SR CNN. However, it is difficult to build accurate correspondences between MODIS and LSR Landsat images due to the effects of atmosphere, weather, terrain and many other complex factors during capturing remote sensing images.

At the same time, accurate reconstruction from LSR Landsat images to Landsat images is hard because of the existence of large spatial resolution gap. Therefore, define the predicted images from NLM CNN and SR CNN as transitional images and further improve them by utilizing available information via a fusion model.

B. Prediction stage

Since the spatial resolutions of transitional images and LSR Landsat images are very similar the temporal change information of transitional images is utilized to predict the LSR Landsat image by applying high pass modulation. Denote the LSR Landsat images down sampled from Landsat image via bicubic interpolation by the fusion of transitional images and the LSR Landsat image is achieved through the following high pass modulation equation

Peak Signal to Noise Ratio (PSNR) is calculated for bicubic interpolation and Super Resolution CNN of each band.

Root Mean Square Error (RMSE) is calculated as the equation given below

Σ Σ (( ) ( ))

RMSE= (5)

where L,ˆare real observed and fusion image. R, C are image height and width

Structural Similarity (SSIM) is calculated as

SSIM= (ˆ)(ˆ)

(ˆ)(ˆ) (6)

where ˆvariances of are predicted and observed images.

ˆis covariance between predicted and observed images.

ˆare mean of predicted and observed images. are small constant to avoid instability, the range is between 0 and 1.

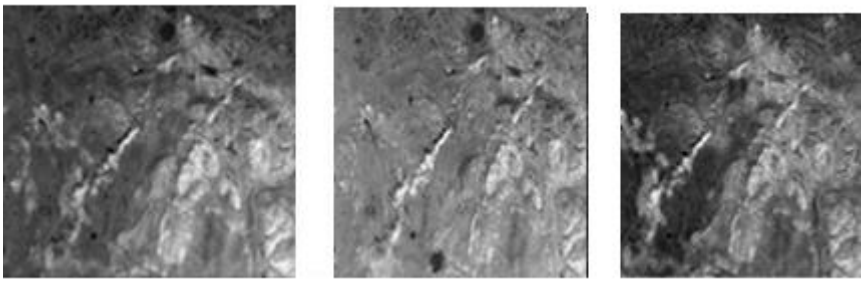
IV. EXPERIMENTAL RESULTS

Remote sensing images with various resolutions can be observed by many earth observing satellites. Extracting the features by illumination and bicubic interpolation and the feature fusion is done by super resolution convolutional neural network. In prediction stage, fused image is compared with other predefined data.

A. INPUT IMAGE

Figure 2 the temporal dynamics mainly associate with crop phenology over a single growing season, but the surrounding agricultural and woodland areas are view the MODIS image.

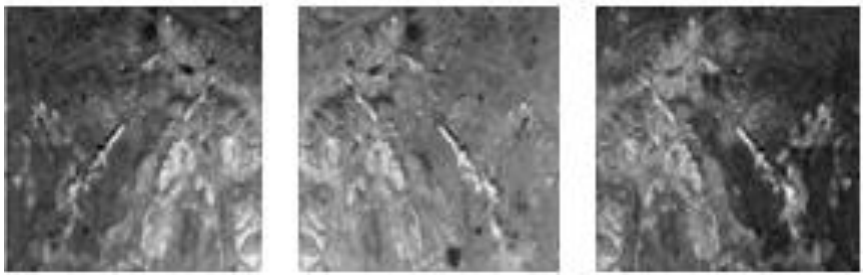
Figure 2 Input image



(a) (b) (c)

Figure 3 Rescale image

Figure 3 the rescaling is frequently used for sorting images to the required size. The fusion of transitional images and the LSR Landsat image from t1 end is achieved through the high pass modulation.

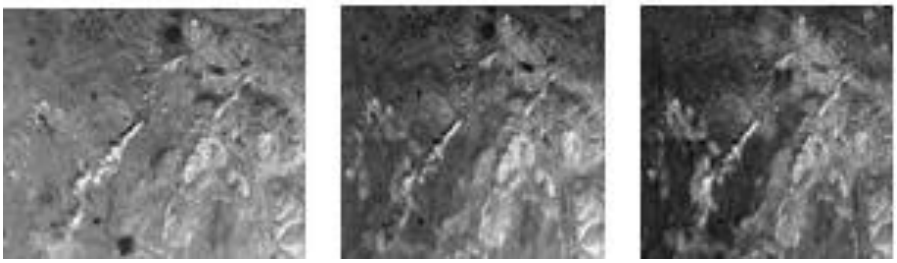


(a) (b) (c)

Figure 4 Bicubic interpolation

Figure 4 the size of this reconstruction filter is k3 × k3. Minimization of the loss between the reconstructed residual images and the corresponding ground truth residual images R.

The red band reconstructed image, is the green band reconstructed image, is the blue band reconstructed image are obtained. This means can add a convolutional layer to reconstruct the residual images.



(a) (b) (c)

Figure 5 SRCNN reconstruction

Figure 5 is SRCNN reconstructed image for red, green and blue band are obtained.

B. RESULTS

A convolutional neural network is a novel spatiotemporal fusion model using deep convolutional neural networks. In the prediction stage, a fusion model consisting of high pass modulation and weighting is adopted to make full use of the available information.

Fusion is a method of combining source images taken from the same scene. A convolutional neural network is used to extract the high frequency details from the two source images.

Figure 6 Fused image

C. PERFORMANCE ANALYSIS

Table .1 Fused image performance analyses

FEATURES	PROPOSED WORK
Structural Similarity (SSIM)	0.914
Peak Signal to Noise Ratio (PSNR)	20.95 dB
Root Mean Square Error (RMSE)	0.022

SSIM values ranges between 0 to 1 and SSIM values 0.93, 0.9 0.91, 0.92 are for good quality reconstruction. PSNR ratio is used as an error measurement between the original and a fused image and the range is between 15 dB to 25 dB and greater than value gives improved image. RMSE values between 0.02 and 0.055 shows that the fused image can relatively predict the data accurately and lesser value shows that it is the improved result.

V. CONCLUSION

Based on deep CNNs, proposed a spatiotemporal fusion method to combine the spatial information of Landsat images and the temporal information of MODIS images. Considering the complex correspondence relationship and the large spatial resolution gap between MODIS and Landsat images, first learn an NLM CNN between MODIS and LSR Landsat images. Then, learn an SR CNN between LSR Landsat and original Landsat images. Via residual learning and back propagation, a five layer NLM CNN and a five layer SR CNN are learned in training stage. In prediction stage, predict the Landsat image on prediction date from two given prior Landsat MODIS image pairs and the corresponding MODIS image. To fully utilize the prior information, define the predicted images from the NLM CNN and SR CNN as transitional images and then adopt a high pass modulation to integrate the information of prior LSR Landsat or original Landsat images.

VI. REFERENCES

[1]. Andrew A. Nicholas, 'Multiscale analysis of relationship between imperviousness and urban tree height using airborne remote sensing', Remote Sensing of Environment, Vol. 113, No. 9, pp. 1988– 1999.

[2]. Beena Matikainen and K. Karila, 'Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating', ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 29, No. 2, pp. 271–285.

[3]. R. Hang and Qingshan Liu, 'Learning Multiscale Deep Features for High-Resolution Satellite Image Scene Classification', IEEE transactions on geoscience and remote sensing, Vol. 56, No. 3, pp. 3421-3567.

[4]. L. Jodriges, Wei Li and G. Wu, 'Hyperspectral Image Classification Using Deep Pixel-Pair Features', IEEE transactions on geoscience and remote sensing, Vol. 112, No. 6, pp. 2914–2926.

[5]. G. Magnus, Frosti Palsson and Boon Giie Lie, 'Multispectral and Hyperspectral Image Fusion Using 3-D-CNN', IEEE geoscience and remote sensing, Vol. 156, No.2, pp. 169–181.

[6]. Ming Aang, Kai Zhang, Wang chang and L. Jiao,'Convolution Structure Sparse Coding for Fusion of Panchromatic and Multispectral Images', IEEE transactions on geoscience and remote sensing, Vol. 54, No. 12, pp. 7135–7148.

[7]. Ming Shang, M. Keiseler and Y. Zhang,'Enhancing Spatio-Temporal Fusion of MODIS and Landsat Data by Incorporating 250 m MODIS Data', IEEE journal of selected topics in applied earth observations and remote sensing, Vol. 44, No. 8, pp. 2207–2218.

[8]. Minji Guo and J. Li, 'An Online Coupled Dictionary Learning Approach for Remote Sensing Image Fusion', IEEE journal of selected topics in applied earth observations and remote sensing, Vol.7, No.4, pp. 4345-4657.

[9]. Muanfeng Shen, K. Jha and X. Meng, 'An Integrated Framework for the Spatio–Temporal–Spectral Fusion of Remote Sensing Images', IEEE transactions on geoscience and remote sensing, Vol. 54, No. 12, pp. 2343-2543.

[10].C. Nuong and Xudong Guan, 'An Object-Based Linear Weight Assignment Fusion Scheme to Improve Classification Accuracy Using Landsat and MODIS Data at the Decision Level', IEEE transactions on geoscience and remote sensing, Vol. 55, No. 12, pp. 4546-4986.

[11].A. Ozg, S. Kwon and S. Liang, 'Merging the MODIS and Landsat Terrestrial Latent Heat Flux Products Using the Multiresolution Tree Method', IEEE transactions on geoscience and remote sensing, Vol. 10, No. 9, pp. 4116–4123.

[12].Pang Xu, J. Dimitrios Marmanis, Lee and M. Datcu, 'Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks', IEEE geoscience and remote sensing letters, Vol. 13, No. 1, pp. 4476–4488.

[13].Penghai, Niraj and K. Cao, 'Spatial and Temporal Image Fusion via Regularized Spatial Unmixing', IEEE Geoscience Remote Sensing, Vol. 5, No. 3, pp. 453–457.

[14].Rhia Xu and H. Shen, 'Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature', Remote Sensing Environment, vol. 113, no. 9, pp. 1874–1885.

[15].X.Tang, Chao Dong and A. Hernado, 'Image super- resolution using Deep convolutional networks', IEEE Trans. Geoscience Remote Sensing, Vol. 50, No. 10, pp. 3707–3716.

[16].Tixiang Yin, J. Asazo and Penghai Xu, 'Spatiotemporal Fusion of Land Surface Temperature Based on a Convolutional Neural Network', IEEE Transaction on Geoscience Remote Sensing, Vol. 51, No. 4, pp. 1883–1896.

[17].Wang Jingan and Q. Cheng, 'Spatiotemporal Fusion with Only Two Remote Sensing Images as Input', IEEE Journal of Selected Topics Applied Earth Observation Remote Sensing, Vol. 7, No. 4, pp. 1284– 1294.

[18].Wei Jingbo, E. Gil and L. Wang, 'Spatiotemporal Fusion of MODIS and Landsat-7 Reflectance Images via Compressed Sensing', IEEE Transaction on geoscience and Remote Sensing, Vol. 7, No. 5, pp. 1792–1805.

[19].Xan Liu, J. Mario and G. Huang, 'Fast and Accurate Spatiotemporal Fusion Based Upon Extreme Learning Machine', IEEE Geoscience and Remote Sensing, Vol. 40, pp. 769–776.

[20].Xiu Bee, Liguo Wang and M. Atkinson, 'Investigating the Influence of Registration Errors on the Patch-Based Spatio-Temporal Fusion Method', IEEE Journal of selected topics in applied earth observations and remote sensing, Vol. 38, No. 2, pp. 295–307.

[21].Yejian Zhou, Sanghan and L. Zhang, 'Optical-and- radar Image Fusion for Dynamic Estimation of Spin Satellites', IEEE Transaction on image processing, Vol. 32, pp. 580–587.

[22].Yhenfeng, Madhi Khaii and J. Cai, 'Remote Sensing Image Fusion with Deep Convolutional Neural Network', IEEE Journal of Selected Topics Applied Earth Observation Remote Sensing, pp. 1097–1105.

[23].Yuanyu Shi and S. Liang, 'A Method for Consistent Estimation of Multiple Land Surface Parameters from MODIS Top-of-Atmosphere Time Series Data', IEEE Transaction on geoscience and Remote Sensing, Vol. 3, pp. 807–814.

[24].Zrsan Batu, Seon and D. Maktav, 'Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey', IEEE Transaction on geoscience and Remote Sensing, Vol. 2, pp. 448–456.

[25].Zu Chen, Mark Wan and W. Shen, 'Evidential Fusion Based Technique for Detecting Landslide Barrier Lakes from Cloud-Covered Remote Sensing Images', IEEE Journal of Selected Topics Applied Earth Observation Remote Sensing, Vol.9, pp. 770–777.