# Satellite-based Spatio-temporal Data Fusion: Current Status and its Implications

**Khaled Hazaymeh, Quazi K. Hassan, and Khan R. Rahaman**

Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary, 2500 University Dr. NW, Calgary, Alberta, Canada T2N 1N4

## Abstract

Satellite image fusion is considered an advanced image processing method. It is used to combine information from two or more satellite images into a single image. It propounds more information and visual interpretation possibilities for future processing. Recently, satellite image fusion methods have been moved from generic to advanced methods that applied statistical or mathematical frames to enhance both the spatial and the temporal resolutions at the same time. The focus of this paper is to review various recently developed spatio-temporal image fusion methods, such as spatio-temporal adaptive fusion methods; unmixing-based fusion methods; and sparse representation-based fusion methods, in order to comprehend the current developments and its implications.

## 1. Introduction

Earth observation satellites provide data of the Earth surface in the form of imagery at different spatial, temporal, and spectral resolutions (Jensen 2007). Usually, high spatial resolution data provide the spatial details of the Erath features, while high temporal resolution data detect the dynamic phenomena (Hilker et al., 2009). Generally, due to technical and cost issues, relatively high spatial resolution images are having low temporal resolution, and vice versa (Zhang 2004; Rocha et al. 2012; Weng et al., 2014). However, many environmental applications (e.g., agriculture, land cover disturbance monitoring, water management, etc.) require both high spatial and high temporal resolution images. Thus, a new area of research has been emerged called spatio-temporal data fusion. Basically, this approach employs statistical or mathematical frameworks to produce high spatial and high temporal satellite images by integrating images from different satellite systems that have similar spectral and orbital characteristics, but with different spatial and temporal resolutions. In this paper, our focus is to present a state of the art of recently developed satellite based spatio-temporal data fusion methods in order to comprehend the current developments and its implications.

## 2. Satellite based spatial-temporal data fusion methods

Spatio-temporal data fusion approach was introduced by Gao et al. (2006) in their effort to generate daily synthetic surface reflectance images at 30 m spatial resolution through the blending of the Landsat (having 30 m spatial resolution at 16 day intervals) and MODIS (having

500 m resolution at daily time interval) surface reflectance images. Since then, several spatio-temporal data fusion methods have been developed (Roy et al., 2008; Zhu et al., 2010; Zhang et al., 2013; Fu et al., 2013, Weng et al., 2014). These methods were introduced a combination of different satellite images, such as: (i) Landsat and MODIS (Gao et al., 2006; Zhu et al., 2010; Fu et al., 2013) (ii) Landsat and MERIS (Zurita-Milla et al., 2009); (iii) HJ-1 and MODIS (Meng et al., 2013); (iv) Landsat and GOES (Wu et al., 2013), and (v) ASTER and MODIS (Liu and Weng 2012). In general, the main idea of the spatio-temporal data fusion approach is to fuse low-temporal/high-spatial resolution data with high-temporal/low-spatial resolution data. This method was applied to different environmental parameters and applications, with major emphasis in: (i) surface reflectance (Gao et al., 2006; Zhang 2013), (ii) normalized difference vegetation index (NDVI) (Meng et al., 2013), (iii) evapotranspiration (Anderson et al., 2010), (iv) urban heat island (Huang et al., 2013), (v) public health (Liu and Weng 2012), and (vi) surface temperature (Weng et al., 2014). These spatio-temporal data fusion methods could be broadly classified into three groups, (i) spatio-temporal adaptive fusion methods, (ii) unmixing-based spatio-temporal fusion methods, and (iii) sparse representation-based spatio-temporal fusion models; and would briefly be discussed in the following sub-sections.

### 2.1 Spatio-temporal adaptive fusion methods

These methods used mathematical functions, such as: summation, subtraction, multiplication, and rationing to perform data fusion technique. Gao et al. (2006) introduced the spatial and temporal adaptive reflectance fusion model (STARFM) for combining surface reflectance data from MODIS and Landsat images. STARFM was an empirical fusion method that utilized the spectral and orbital similarities between Landsat and MODIS, and combined them to produce synthetic Landsat images at the temporal resolution of MODIS (i.e., daily or 8-day scale). It consisted of three main steps as:

i. selecting spectrally similar pixels (i.e., having similar reflectance) within a moving window of interest using Landsat images;
ii. determining a weighting factor as a function of both Landsat and MODIS images; and
iii. generating synthetic Landsat images at time two [$synth\text{-}L(t_2)$] by multiplying the weighting factor with the sum of difference between two MODIS images taken at two different times [$M(t_2) – M(t_1)$] and Landsat image taken at time one [$L(t_1)$].

Though the STARFM method predicted reasonable synthetic Landsat images in comparison with reference Landsat images; there were three major limitations: (i) the window size was variable depending on the study area, and needed to be adjusted at each run; (ii) the existence of spectrally similar pixels within the moving window might not be always found; and (iii) it was not applicable over heterogeneous landscapes. In order to address the issue of heterogeneous land cover, Zhu et al. (2010) proposed the Enhanced STARFM (ESTARFM). In this case, they used two pairs of Landsat and MODIS images taken in two different dates such as $(t_1)$ and $(t_3)$ in order to predict synthetic Landsat image at an intermediate time *($t_2$)*. The enhancement included the determination of a conversion coefficient for each spectrally similar pixel by performing linear regression analysis between Landsat and MODIS images [i.e., $L(t_1)$ with $M(t_1)$] and [$L(t_3)$ with $M(t_3)$] prior to generate the synthetic Landsat image. Although the method successfully predicted synthetic Landsat images, it was not only computationally expensive but also had the same problem of the moving window and the spectrally similar pixel issues like the original STARFM. To address the issue of selecting spectrally similar pixels, Fu et al. (2013) modified ESTARFM (mESTARFM). They applied two conditions to the candidate pixels, as: (i) the candidate pixels should be less than or equal to the standard deviation threshold of the moving window, and (ii) had the same land cover type as the central pixel in the moving window. This proposed method improved the accuracy of the predicted image in comparison to that of ESTARFM-produced synthetic images. Despite, it still had other problems such as, long computation time and the inapplicability for near real time applications. In another study, Meng et al. (2013) introduced the spatial and temporal adaptive vegetation index fusion model (STAVFM); where they improved the weighting function of STARFM by defining a time window according to the temporal variation of crops (in the event the method applied over agriculture-dominant land cover types). Although the method improved the predicted normalized difference vegetation index (NDVI) images; however, it failed to address other two major STARFM limitations. In other applications, researchers implemented STARFM in predicting synthetic surface temperature (*Ts*) (Liu and Weng 2012) and evapotranspiration (Anderson et al., 2010). However, the direct implementation of STARFM in these applications revealed severe errors due to the STARFM limitations (heterogeneity, window size, and spectrally similar pixels). Thus, Weng et al. (2014) incorporated the annual temperature cycle and a linear spectral mixing analysis within the original STARFM model and proposed the spatio-temporal adaptive data fusion algorithm for temperature (SADFAT) mapping. Although the predicted *Ts* images showed good agreements with actual/reference *Ts* images, the method required adjusting the size of the moving window and the number of land cove types each time before the model implementation (Weng et al., 2014). Huang et al. (2013) and Wu et al. (2013) applied a bilateral-based filter and a variation-based filter respectively to calculate the predicted value of the *Ts* within the moving window instead of using the original filter-based method of STARFM. However, in both of these studies, *Ts* outliers within the moving window of interest was affected by the predicted images and produced uncertain synthetic *Ts* images.

### 2.2 Unmixing-based spatio-temporal fusion methods

The basis of unmixing-based fusion methods was the employment of a classification technique in generating synthetic image having both high spatial and temporal resolutions. For instance, Zurita-Milla et al. (2009) used a linear unmixing model to un-mix MERIS time-series images acquired at 300 m spatial resolution using a land use database with a 25 m spatial resolution. The approach was based on assigning the unmixed signals to the corresponding land-use class presented in the central pixel of $k*k$ MERIS neighborhood. Though the results were promising in monitoring vegetation dynamics at Landsat-like spatial (~25 m) and MERIS-like spectral and temporal ~3 days) resolution; however, the accuracy of this method was highly dependent on the quality and the availability of the land use database. In another study, Zhang et al. (2013) further intensified the ESTDFM proposed by Zhu et al. (2010) by incorporating ISODATA classification techniques. This particular approach had four main steps:

i. classifying Landsat images using a patch–based ISODATA classification technique and calculating the abundance of end members within a moving window;
ii. unmixing of the three MODIS images using those end members;
iii. predicting two synthetic images by calculating the sum of $L(t_1)$ and $L(t_3)$ with the corresponding difference images of unmixed MODIS images; and
iv. generating the final synthetic image by weighting the two predicted images produced in the previous step.

This method eliminated the requirement of a high spatial resolution land use map as the case in Zurita-Milla et al. (2009). However, it would have the same problems and issues of ESTARFM proposed by Zhu et al. (2010).

### 2.3 Sparse representation-based spatio-temporal fusion methods

Huang and Song (2012) introduced the concept of the sparse representation-based spatio-temporal reflectance fusion model (SPSTFM). They used MODIS images taken at 3 times [$M(t_1)$, $M(t_2)$, and $M(t_3)$] and two Landsat images taken at 2 times [$L(t_1)$, and $L(t_3)$] to generate *synth*-$L(t_2)$. The method consisted of four steps:

i. enhancing the three MODIS images to the equivalent spatial resolution of Landsat image (i.e., 30 m) through a sparse representation technique;
ii. building a dictionary pair (an image processing method) using the two counterpart MODIS and Landsat images [$M(t_1)$ and $L(t_1)$; $M(t_3)$ and $L(t_3)$] and the other MODIS image [$M(t_2)$];
iii. predicting the difference image of Landsat data using the learned dictionary pair; and
iv. reconstructing the predicted Landsat image using different weighting parameters.

This method provided better results (based on $R^2$ and RMSE) when compared with the STARFM-based fusion methods; however, it was computationally expensive and impractical for near real time applications. Song and Huang (2013) improved the SPSTFM method by employing a sparse representation procedure that include dictionary training and spars coding, in order to enhance the spatial resolution of MODIS images (known as transition images) to the spatial resolution of Landsat image. Then two transition images taken at $t_1$ and $t_2$ and one Landsat image [$L(t_1)$] was used to predict the synthetic Landsat image at time 2 ($t_2$) by using a high pass modulation technique. This method used only three input images to predict the synthetic image. Thus, it eliminated one of the SPSTFM problems, however it required relatively longer processing time due to the dictionary training.

## 3.0 Concluding remarks

In this paper, we reviewed three different groups of image fusion methods that were used for enhancing the spatio-temporal resolutions of satellite images. These methods were: spatio-temporal adaptive fusion methods, unmixing-based fusion methods, and sparse representation-based fusion methods. Actually, it was difficult to point out the most reliable method, as the comparisons among them showed that there were no superior one over the others in all circumstances. It was found that the selection of an appropriate method would depend on many factors including data availability and their quality, type of application and area of interests, and landscape properties. In summary, it could be stated that satellite image fusion methods started to move from generic to advanced approaches considering sensor, context, application and other parameters in the process. Discussion from this paper would also open up windows for researchers to conduct more studies with different methods to delineate suitability depending on their area of studies.

## References


Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., d'Urso, G., Pimstein, a., Gao, F., 2010. Mapping daily evapotranspiration at field to global scales using geostationary and polar orbiting satellite imagery. Hydrology and Earth System Sciences Discussions 7, 5957–5990.

Fu, D., Chen, B., Wang, J., Zhu, X., Hilker, T., 2013. An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. Remote Sensing 5, 6346–6360.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Transactions on Geoscience and Remote Sensing 44, 2207–2218.

Hilker, T., Wulder, M.A., Coops, N.C., Seitz, N., White, J.C., Gao, F., Masek, J.G., Stenhouse, G., 2009. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. Remote Sensing of Environment 113, 1988–1999.

Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Transactions on Geoscience and Remote Sensing 50, 3707–3716.

Huang, B., Wang, J., Song, H., Fu, D., Wong, K., 2013. Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring. IEEE Geoscience and Remote Sensing Letters 10, 1011–1015.

Jensen, J.R., 2007. Remote Sensing of the Environment: An Earth Resource Perspective. Second ed. Pearson Prentice Hall. New Jersey.

Liu, H., Weng, Q., 2012. Enhancing temporal resolution of satellite imagery for public health studies: A case study of West Nile Virus outbreak in Los Angeles in 2007. Remote Sensing of Environment 117, 57–71.

Meng, J., Du, X., Wu, B., 2013. Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation. International Journal of Digital Earth 6, 203–218.

Rocha, J., Perdigão, A., Melo, R., Henriques, C., 2012. "Remote sensing based crop coefficients for water management in agriculture". In: Sustainable Development - Authoritative and Leading Edge Content for Environmental Management. Editor: Sime Curkovic, 588 pages, InTech Europe, Croatia.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sensing of Environment 112, 3112–3130.

Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. IEEE Transactions on Geoscince and Remote Sensing 51, 1883–1896.

Weng, Q., Fu, P., Gao, F., 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sensing of Environment 145, 55–67.

Wu, P., Shen, H., Ai, T., Liu, Y., 2013. Land-surface temperature retrieval at high spatial and temporal resolutions based on multi-sensor fusion. International Journal of Digital Earth 6, 113–133.

Zhang, W., Li, A., Jin, H., Bian, J., Zhang, Z., Lei, G., Qin, Z., Huang, C., 2013. An enhanced spatial and temporal data fusion model for fusing Landsat and MODIS surface reflectance to generate high temporal Landsat-like data. Remote Sensing 5, 5346–5368.

Zhang, Y., 2004. Understanding image fusion. Photogrammetric Engineering and Remote Sensing 70, 657–661.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sensing of Environment 114, 2610–2623.

Zurita-Milla, R., Kaiser, G., Clevers, J.G.P.W., Schneider, W., Schaepman, M.E., 2009. Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. Remote Sensing of Environment 113, 1874–1885.