Filling the data gaps in Landsat 7 Images using Texture Analysis

M. Srinivas[1, 2] and P. Rama Chandra Prasad[2]

[1.] CVR college of Engineering, Vastunagar, Mangalpalli, R.R. District.

e-mail: msrinivasremote1975@gmail.com

[2.] Lab for Spatial Informatics, International Institute of Information Technology, Hyderabad.

e-mail: rcprasad@iiit.ac.in

Abstract: Missing data in satellite images is a known phenomenon in remote sensing. There are various reasons for these gaps in satellite images e.g. shadowed area for SAR data sets, cloud coverage for optical imagery and instrument errors such as Scan Line Corrector(SLC-off) failure. The Scan Line Corrector (SLC) of Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor failed permanently causing around 23% of pixels not scanned in each scene and this failure has seriously limited the scientific applications of ETM+ data. Reconstructing the gap regions is an important issue in remote sensing. To address this issue, we have proposed an idea to find the homogeneous and heterogeneous regions of ETM+ image using Morphological operations and replace the unscanned pixel value by Neighborhood similar pixel threshold based Local Binary Pattern value. We have applied padding approach with additional rows and columns for unscanned mixels at borders to replace with the corresponding NLBP value. It is observed that our results are promising to fill the data gaps in both homogeneous and heterogeneous regions.

Key words: Scan Line Corrector, Local Binary Pattern, Landsat 7

1. Introduction: A Series of Landsat satellites has produced enormous amount of data to map and monitor the land surface[1, 2, 3, 4, 5, 6]. The Landsat satellites include the Landsat 5 Thematic Mapper (TM), the Landsat 7 Enhanced Thematic Mapper Plus ETM+), and the Landsat 1–5 Multispectral Scanners(MSS). From the year of launch in 1972, Landsat satellites have provided data with large coverage of 30m – 28.5m resolution for TM and ETM+ and 80m for MSS sensors. On May 31 2003, the scan line corrector (SLC) landsat 7 ETM+ sensor has failed permanently. Due to this the images obtained by SLC showed wedge-shaped gaps ranging from a single pixel in width near the image-nadir to about 12 pixels towards the edges of the scene. About 23% of the information is lost due to SLC failure and resulting in a great obstacle for ETM+ sensor applications[7, 8]. After the SLC failure has occurred, the images acquired are called as SLC-off images and the images before the failure are SLC-on images. As an alternative, a joint collaboration of United States Geological Survey/National Aeronautics and Space Administration (USGS/NASA) Landsat team has developed several methods to fill the data gaps in Landsat images. One among these developed methods is local linear histogram-matching method. Rescaling function is derived from the local linear histogram which in turn is calculated from the sliding window of each moving pixel in SLC-off image. The rescaling function derived from SLC-ff image will be used to convert in-to the radiometric values to fill the gaps and the transformed data is used to fill the gap of that scene. Local linear histogram method is a simple method to implement and it resolves many missing data problems for the images of high quality[9]. Roy et.al. (2008) has proposed another approach to estimate the reflectance of unscanned pixels from MODIS data to fill the gaps in high quality data[10]. To fill the gaps in Landsat 7 ETM+ images multi-scale segmentation approach was developed [11]. Multi-scale segmentation approach was used for land cover mapping and visual assessment[12]. krigging or co-krigging techniques which are Geostatistics based methods are also employed to interpolate the missing pixels in the SLC-off imagery[13, 14].

All the methods described above have certain limitations restricted to particular applications. In the case of local linear histogram method[10] it will produce good results for homogenous areas like forests but not for heterogeneous areas where the size of object is smaller than window size. Sensors which do not belong to Landsat are limited by spectral and spatial resolutions issues, for example the data obtained from MODIS which is similar to ETM+ in terms of spectral bands are not advantageous due to their coarser spectral resolution in obtaining the reflectance values for filling the data gaps. For narrow or smaller objects like roads, streams multi-scale segmentation approach is not suitable at the pixel level which results in lower prediction accuracy[11]. Geostatistical interpolation has a drawback of predicting the pixel value for smaller objects and they are computationally expensive which is limiting for larger applications[12, 13].

To overcome these shortcomings, the objective of the current study is to demonstrate the application of simple and effective method to fill the data gaps in Landsat 7 ETM+ SLC-off imagery. This method will identify the local regions and it is based on the threshold of neighborhood similar pixel interpolator which will be called as Neighborhood similar pixel threshold based Local Binary Pattern (NLBP) approach to fill the gaps of missing pixels especially for heterogeneous areas. In this paper, we will first describe the algorithm proposed and later it will be demonstrated with simulated results of the filled gaps with the Local Binary pattern approach (LBP) for SLC-off images.

2. Algorithm development:

The neighboring pixels which belong to the same class will have similar spectral characteristics and temporal patterns will share the similar values of neighboring pixel values to fill the gaps. For the experiments, we considered the Landsat 7 ETM+ SLC-off data of Hyderabad region from USGS Earth explorer acquired on February 20, 2016. To fill the data gaps, we have proposed NLBP algorithm with the basic idea originated from Local Binary Pattern (LBP) approach for texture analysis. In this section, we will start with the morphological operations to identify the homogeneous and heterogeneous regions based on the size and shape of the structural element followed with the construction of algorithm for NLBP.

Morphological operations are a set of operations that process images based on shapes. These operations are evaluated to find the similarities between unscanned pixels and scanned pixels which belong to same homogenous region. They apply a structuring element to an input image and generate an output image. The two basic operations of morphological process are Erosion and Dilation[15]. The steps required for Erosion and Dilation are given below:

2.1. Algorithm: Erosion.

1. A kernel (a matrix of odd size (3, 5, 7) is convolved with the image.
2. A pixel in the original image (either 1 or 0) will be considered 1 only if all the pixels under the kernel are 1, otherwise it is eroded (made to zero).
3. Thus all the pixels near boundary will be discarded depending upon the size of kernel. So the thickness or size of the foreground object decreases or simply white region decreases in the image.

2.2 Algorithm: Dilation.

1. A kernel (a matrix of odd size (3, 5, 7) is convolved with the image
2. A pixel element in the original image is '1' if at least one pixel under the Kernel is '1'.
3. It increases the white region in the image or size of foreground object increases.

2.3 Basics of Local Binary Pattern (LBP) operator

Oja et.al. has proposed a non-parametric LBP operator based on local neighborhood around a pixel to represent the texture[14]. Further, rotation invariant and uniform local binary patterns were also proposed as an extension for multi-resolution gray-scale texture classification. For an image, LBP code with P equally spaced pixels with radius R from the centre pixels is computed for the unscanned pixels are given below:

$$\mathrm{LBP}_{P,R}=\sum_{i=0}^{p-1} S(x)2^{p} \quad \text{-------------} \quad (1)$$

$$x=G_P-G_C \quad \text{------------------} \quad (2)$$

In (2) G_P and G_C are the gray values of the neighborhood pixel and centre pixel respectively. The LBP code is generated for all the unscanned pixels of the ETM+ image. Finally, occurrences of LBP codes of unscanned pixels are counted to form a LBP histogram which is used to represent the texture.

2.4 Neighborhood similar pixel threshold based LBP

In Neighborhood similar pixel threshold based LBP, the binary pattern is based on threshold which is computed using bell shaped membership function. The membership value is assigned to difference of neighboring pixel and centre pixel. The mean of memberships of differences is taken as threshold. The algorithm for computing NLBP was implemented to extract the features of ETM+ image. The inputs I, n, and R refer to an ETM+ image, neighborhood size, and radius respectively. $Diff_j$ is the difference between j^{th} neighbour p_j and centre pixel p_c. Norm is a mean of differences, $Diff_j$ and μ_j is the bell shaped membership function. The steps required to fill the data gaps are given below:

<u>**Algorithm:**</u> Neighborhood similar pixel threshold based LBP

1. While not end in ETM+ image
2. Get Window W_I from ETM+ of the radius R chosen
3. Compute $Diff_j = p_j - p_c$
4. Compute the sum of differences, call it as $Diff_j$ for 'n' iterations
5. Calculate the Norm of the differences
6. Obtain the ratio $\mu_j = \frac{\text{Diff}_j}{\text{Norm}}$
7. Obtain the threshold with the mean of ratios obtained from step 6
8. Calculate the NLBP similar to the LBP procedure derived from (1)
9. End while
10. Explore with different values of R for better features.

After evaluating NLBP of unscanned pixels, morphological operations are again applied to process the image at pixel borders, because at borders the unscanned mixels will be replaced with the corresponding value of NLBP which will be evaluated based on more number of pixels occupied in that homogeneous region.

3. Data preparation:

The study region is located in Hyderabad of Telangana region around 17.3° N and 78.4° W and is covered by World Reference system 2 path 114 and Row 68. We selected entire study area to differentiate homogeneous and heterogeneous regions of unscanned pixels with our proposed algorithm. Majority of the area include forest, water and arable land and the land covers both "green" and "nongreen" crops. All the 8-band images were collected with mask for each band is considered to fill the gaps in the image.

4. Results and Discussion:

To fill the data gaps we have considered the SLC-off stacked image with the layers of bands 3, 4 and 5 shown in Figure 4(a). There were a total of 125000 unscanned pixels in the image.

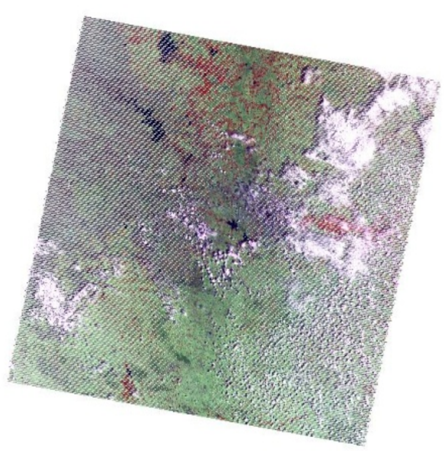

Figure 4(a). SLC-off image with the combinations of band 3, 4 and 5 of Hyderabad region

Data Gap filling was done with neighborhood similar pixel threshold based LBP approach described in Algorithm 2.4. We compared the local linear histogram method described by the USGS [11] and with our algorithmic approach for a study area of 15km X 15 km. which are shown in Figure 4(b) and Figure 4(c).

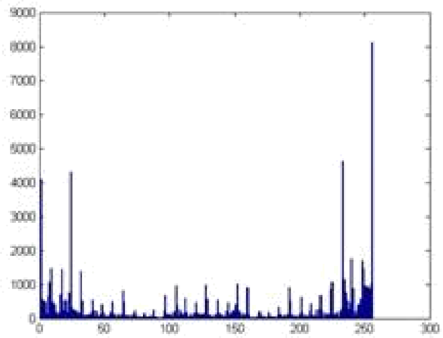

Figure 4(b). Local linear Histogram

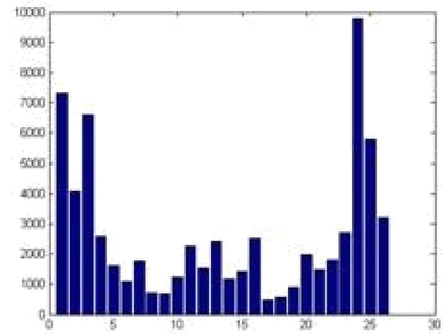

Figure 4. (c). NLBP Histogram of proposed algorithm

Local Linear Histogram proposed by USGS [10] has a drawback of identifying the unscanned pixels at the borders. That's why there are gaps in the histogram due to mixels shown in Figure 4(b). But these gaps are filled with our proposed algorithm which was shown in Figure 4(c). From literature it is observed that there are different interpolation approaches to fill the data gaps and we felt that conducting similar pixel based comparisons between Neighborhood similar pixel interpolation and the multi-scale segmentation approach[12] or the geostatistics based methods[13,14] was statistically problematic and beyond the scope of the study.

Initially, we have applied morphological operations to know the similarities between scanned pixels and unscanned pixels. By using the erosion and dilations operations of morphological process we have identified homogeneous and heterogeneous regions based on their shapes. For each unscanned pixel belonging to homogeneous and heterogeneous regions of that class, we have evaluated the Neighborhood similar pixel threshold based Local Binary Pattern value described in Algorithm 2.4. The unscanned pixel value is replaced with the calculated NLBP value. During the process, we left the mixels belonging to the border regions because the mixel will share common neighborhood similar pixel threshold based LBP value between two homogeneous regions.

To fill the data gaps for the pixels at borders the morphological functions are evaluated and we have assigned a value to these undefined pixels, as if the functions had padded the image with additional rows and columns. The value of these padding pixels varies for dilation and erosion operations. The following table describes the padding rules for dilation and erosion are given below

Operation	Rule
Dilation	Pixels beyond the image border are assigned the minimum value of NLBP either from region 1 or region 2, if the numbers of pixels from two regions are same.
Erosion	Pixels beyond the image border are assigned the *maximum* value of NLBP if the number of pixels in region 1 is more compared to region 2 or vice versa

Table 4.1 Padding rules for Mixels

After padding with additional rows and columns around the borders in the image the NLBP is calculated for mixels similar to the procedure described in Algorithm 2.4. Figure 4(d) show the corresponding output image.

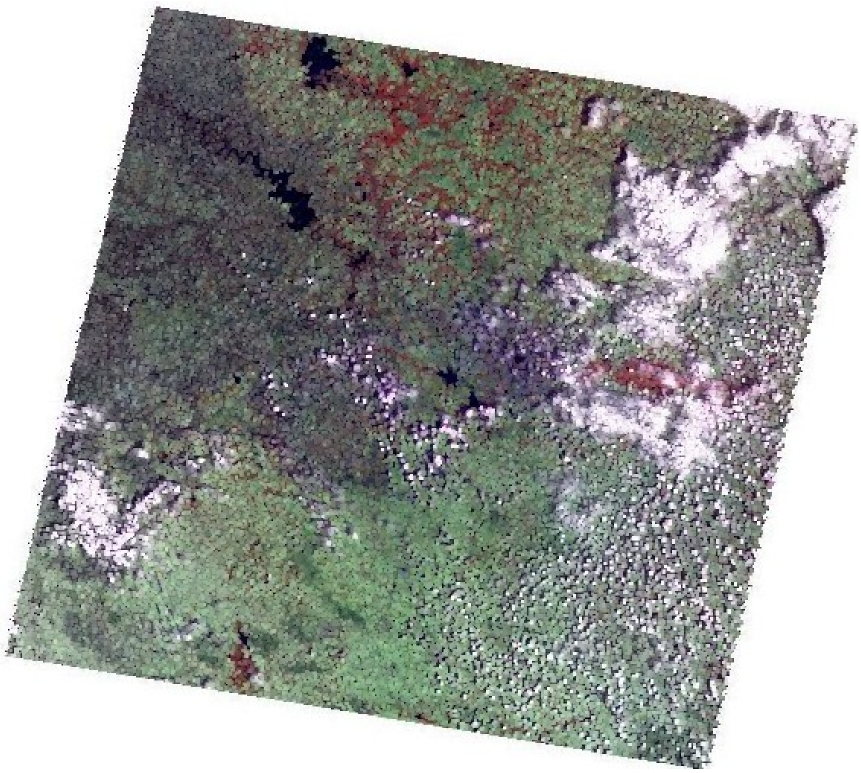

Figure 4(d) corrected image with the proposed NLBP algorithm

5. Conclusions:

Despite of the SLC failure the quality of the radiometry and geometry of the Landsat 7 ETM+ data is still important for many applications. It is very much needed to develop techniques for filling the un-scanned gaps in the SLC-off imagery. Unfortunately existing methods has certain limitations. This paper proposes a new and effective method called as Neighborhood similar pixel threshold based LBP to fill the data gaps in Landsat 7 ETM+ SLC-off imagery. The NLBP can restore all the unscanned pixels accurately and worked well even for heterogeneous regions. The major improvement of NLBP approach is that it makes better use of scanned pixels belonging to that class without loss of any relevant information to fill the unscanned pixels. To differentiate the similarity between scanned and unscanned pixels morphological operations are applied to find the homogeneous and heterogeneous regions based on shape and size of the patch in that region. We compared the Local Linear histogram proposed by USGS and our algorithmic approach and it is concluded that the mixels at the borders will be clearly identified and there is an improvement in the occupancy of unscanned pixels. The NLBP calculated for each unscanned pixel of homogeneous and heterogeneous region shows that there is a better improvement of filling the gaps in Landsat 7 ETM+ images. Further the work can be extended in removing the uncertainty of mixels for obtaining a better accuracy of classification which is important for land cover and mapping applications.

6. References:

1. Arvidson, T., Goward, S., Gasch, J., & Williams, D. (2006). Landsat-7 long-term acquisition plan: Development and validation. Photogrammetric Engineering and Remote Sensing, 72, 1137−1146.

2. Byrne, G. F., Crapper, P. F., & Mayo, K. K. (1980). Monitoring land-cover change by principal component analysis of multi-temporal Landsat data. Remote Sensing of Environment, 10, 175−184.
3. Bédard, F., Reichert, G., Dobbins, R., & Trépanier, I. (2008). Evaluation of segment-based gap-filled Landsat ETM+ SLC-off satellite data for land cover classification in southern Saskatchewan, Canada. International Journal of Remote Sensing, 29, 2041−2054.
4. Chander, G., Xiong, X., Choi, T., & Angal, A. (2010). Monitoring on-orbit calibration stability of the Terra MODIS and Landsat 7 ETM+ sensors using pseudo- Invariant test sites. Remote Sensing of Environment, 114, 925−939.
5. Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 sensors. Remote Sensing of Environment, 113, 893−903.
6. Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. Bioscience, 54, 535−545.
7. Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37, 35−46.
8. 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.
9. Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., & Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. Remote Sensing of Environment, 112, 2495−2513.
10. Ju, J. C., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. Remote Sensing of Environment, 112, 1196−1211.
11. Maxwell, S. K., Schmidt, G. L., & Storey, J. C. (2007). A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images. International Journal of Remote Sensing, 28, 5339−5356.
12. Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., et al. (2010). Webenabled Landsat data (WELD): Landsat ETM+ composited mosaics of The conterminous United States. Remote Sensing of Environment, 114, 35−49.
13. Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., et al. (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.
14. T. Ojala and Matti Pietikainen, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on Pattern Analysis and Machine Intelligence. 24:87, 2002.
15. Bernhard Burgeth,Martin Welk,Christian Feddern,and Joac him Weicker "Morphological Operations on Matrix-Valued Images".