A Spatio–Temporal Pixel-Swapping Algorithm for Subpixel Land Cover Mapping

Yong Xu, and Bo Huang, Associate Member, IEEE

Abstract—The aim of this letter is to present a spatio–temporal pixel-swapping algorithm (STPSA), based on conventional pixel-swapping algorithms (PSAs), in which both spatial and temporal contextual information from previous land cover maps or observed samples are well integrated and utilized to improve subpixel mapping accuracy. Unlike conventional pixel-swapping algorithms, STPSA is capable of utilizing prior information, which was previously ignored, to predict the attractiveness based on pairs of subpixels. This algorithm involves three main steps and operates in an iterative manner: 1) it predicts the maximum and minimum attractiveness of each pair of pixels; 2) ranks the swapping scores based on the attractiveness of all the pairs; and 3) swaps the locations of the pair of pixels with a maximum score to increase the objective function. Experiments with actual satellite images have demonstrated that the proposed algorithm performs better than other algorithms. In comparison, the proposed STPSA’s better performance is due to the fact that prior information used in other algorithms is restricted to a percentage level rather than the real subpixel level.

Index Terms—Land cover mapping, pixel swapping, spatial and temporal information, subpixel.

I. INTRODUCTION

SATellite data have proved to be useful for land cover mapping on different spatial scales [1]. Due to the problem of mixed pixels, particularly in coarse-resolution satellite data, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) with a 1-km spatial resolution, securing a high-quality land cover map from satellite data remains a challenge [2]. Correspondingly, a variety of soft classification techniques utilizing mixed pixels are proposed to estimate proportions of each type of land cover within a pixel [3]. Compared with the hard classification (HC) method, soft classification techniques provide the proportions of each individual class within a pixel, rendering them more informative and useful in applications related to predicting or detecting small features within a pixel [4].

Even though soft classification algorithms can obtain the proportions of each individual class within a pixel, they fail to determine the spatial distribution of each class [5]. Therefore, subpixel mapping techniques [3], or super-resolution mapping techniques [5], have been proposed to predict the spatial distribution of subpixels within a coarse pixel by using the proportions of each individual class.

Most of the existing subpixel mapping approaches explore spatial dependence or spatial correlation between pixels or subpixels, such as linear optimization techniques [6], genetic algorithms [7], neural networks [8], interpolation based algorithms [9], super-resolution based approaches [5], and pixel-swapping algorithms (PSA) [3]. PSA [3] is a widely used subpixel mapping algorithm because of its simplicity and fast computation. An improved pixel-swapping algorithm was also developed by Makido et al. [10] to handle multiple land cover classes.

Other than spatial correlation information, usage of spatial pattern or texture information is another consideration. Geostatistical techniques have been well exploited and employed for subpixel mapping [5]. In these approaches, spatial pattern or texture information is firstly learned and captured according to prior training data, and then the trained pattern or texture is applied to coarse resolution data for subpixel mapping. Using a similar idea, spatial pattern or texture was learned using the Hopfield neural network [8], [11] or Markov random field [12]. Some other auxiliary information, such as multi-spectral information [12], [13], was also utilized for improving the accuracy of subpixel mapping.

However, prior information in previous land cover maps and some observed land cover samples have been significantly ignored in most of the proposed approaches. Ling et al. [14] attempted the use of prior land cover percentages to develop a subpixel land cover change mapping (SLCCM) algorithm, with which the original subpixel mapping result of PSA was improved. In the subpixel mapping method devised by Boucher and Kyriakidis [5], observed sample data were used, but they are restricted to the calculation of the statistical parameters, which are very computation intensive. Consequently, a spatio–temporal pixel-swapping algorithm (STPSA), based on conventional PSAs is proposed in this letter, in which spatial and temporal contextual information in previous land cover maps, or observed samples, are well integrated and utilized.

II. METHODOLOGY

This section presents the detailed process of the STPSA by using previous land cover maps and land cover proportions as input.

The proposed approach is implemented based on the coarse land cover proportions of each pixel. Initially, labels of subpixels are randomly allocated maintaining the proportions.
After random initialization, the spatial arrangement of the labels of the subpixels is updated via the following three basic steps.

First, the attractiveness in STPSA is predicted based on pairs of pixels rather than pixel-by-pixel in the conventional PSA [3]. As a pair of subpixels \( i \) and \( j \) with the second-order nearest neighbors (as shown in Fig. 1), the maximum and minimum attractiveness of this pair of subpixels can be calculated as follows:

\[
A_{\text{max}}(z(x_i), z(x_j)) = \sum_{k \in C_i} \lambda_{ik} I(z(x_j), z(x_k)) + \sum_{k \in C_j} \lambda_{jk} I(z(x_i), z(x_k)) \quad (1)
\]

\[
A_{\text{min}}(z(x_i), z(x_j)) = \sum_{k \in C_i} \lambda_{ik} I(z(x_i), z(x_k)) + \sum_{k \in C_j} \lambda_{jk} I(z(x_j), z(x_k)) \quad (2)
\]

where \( z(x_i) \) and \( z(x_j) \) denote labels of subpixels at locations \( x_i \) and \( x_j \); \( C_i \) and \( C_j \) reflect neighborhoods of pixels \( i \) and \( j \); \( I(z(x_i), z(x_j)) \) is an indicator function; \( \lambda \) is a distance decay-based weight function.

The indicator function \( I(x_i, x_j) \) can be predicted as follows:

\[
I(z(x_i), z(x_j)) = \begin{cases} 
1, & z(x_i) = z(x_j) \\
0, & \text{otherwise}.
\end{cases} \quad (3)
\]

The distance decay-based weight function \( \lambda \) [15] is estimated as

\[
\lambda_{ij} = \exp\left( -\frac{-h_{ij}}{a} \right) * \exp\left( -\frac{-t_{ij}}{b} \right) \quad (4)
\]

where \( h_{ij} \) and \( t_{ij} \) are parameters of the spatial and temporal distances between locations \( x_i \) and \( x_j \) of pixels \( i \) and \( j \) for which the attractiveness is desired, while \( a \) and \( b \) are spatial and temporal bandwidths in this exponential model to construct the weight.

While several alternative distance decay-based weight functions, e.g., the inverse distance weighting (IDW) function or the Gaussian function [16], can be chosen here, the nearest neighbor function with an equivalent weight \( \lambda \) is adopted in this study for its simplicity and efficiency.

In STPSA, neighbors for attractiveness prediction are extended from the spatial extent in the original PSA algorithm to spatial and temporal extents. Fig. 2 shows a second-order spatial and temporal neighborhood system, in which each internal pixel has eight spatial neighbors and nine temporal neighbors at each timeslot. Pixel \( i \) at time \( t_2 \), for example, not only has eight nearest spatial neighbors, but also nine temporal neighbors at the previous time \( t_1 \), and this spatial and temporal information is included under the constructed neighborhood system. The neighborhood system can also be extended into a higher-order form and includes more spatial or temporal contextual information.

Under the above defined neighborhood system, prior temporal information is employed in clique \( C \) for the attractiveness calculation of each pair of subpixels, and finally affects the subpixel mapping result.

Second, the swapping attractiveness based on pairs of subpixels can be scored by using the following objective function:

\[
A(z(x_i), z(x_j)) = A_{\text{max}}(z(x_i), z(x_j)) - A_{\text{min}}(z(x_i), z(x_j)) \quad (5)
\]

where \( A_{\text{max}} \) and \( A_{\text{min}} \) represent the maximum and minimum attractiveness of a pair of subpixels.

As a pair of subpixels at locations \( x_i \) and \( x_j \) without prior temporal information, for example, with equal weights to all second-order neighbors, the maximum and minimum attractiveness can be calculated with (1) and (2), and are 8 and 6, respectively. Furthermore, the swapping attractiveness is 2 from (5).

Third, locations of the pair of subpixels with the maximum swapping attractiveness are swapped. Fig. 1(b) shows the result after locations \( i \) and \( j \) are swapped as this pair of subpixels has the maximum swapping attractiveness in this area.

The process of the STPSA includes a number of iterations of the above three steps, which results in a solution.

### III. Case Study

The STPSA was tested using the actual data in the study area (centered at 22.66° N, 114.19° E) of Shenzhen, China. Shenzhen was chosen as the study area based on two considerations: 1) it has experienced a high percentage of land use/land cover change during the past several decades and 2) it is a typical heterogeneous region with various types of land cover. These features make it highly suitable for the validation of the proposed algorithm.
Fig. 3. Tests of subpixel mapping algorithm applied to a degraded Landsat image at the study area of Shenzhen with 300 m spatial resolution (or 300 m for short). (a) Original image at year 2002 (30 m). (b) Degraded image for the algorithm (300 m). (c) Proportions of forest (300 m). (d) Proportions of vegetation (300 m). (e) Proportions of built-up area (300 m). (f) Land cover map at year 2000. (g) Subpixel mapping with bicubic interpolation (30 m). (h) Subpixel mapping with PSA (30 m). (i) Subpixel mapping with SLCCM (30 m). (j) Subpixel mapping with our enhanced algorithm using prior information (30 m). (k) HC for comparisons (300 m).

TABLE I

<table>
<thead>
<tr>
<th>Method</th>
<th>Unchanged area</th>
<th>Changed area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>HC</td>
<td>OA (%)</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td>Kappa (%)</td>
<td>88.3</td>
</tr>
<tr>
<td>BI</td>
<td>OA (%)</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td>Kappa (%)</td>
<td>83.7</td>
</tr>
<tr>
<td>PSA</td>
<td>OA (%)</td>
<td>88.5</td>
</tr>
<tr>
<td></td>
<td>Kappa (%)</td>
<td>81.3</td>
</tr>
<tr>
<td>SLCCM</td>
<td>OA (%)</td>
<td>92.3</td>
</tr>
<tr>
<td></td>
<td>Kappa (%)</td>
<td>87.5</td>
</tr>
<tr>
<td>STPSA</td>
<td>OA (%)</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td>Kappa (%)</td>
<td>92.9</td>
</tr>
</tbody>
</table>

Acquired data in this study area includes a Landsat satellite image with 30 m spatial resolution on November 7, 2002 and the land cover survey data at year 2000. As Fig. 3(a) shows, a subimage of 250 pixels by 250 pixels (7.5 km by 7.5 km) is extracted from the original image. It represents a mixed landscape with natural land cover elements (water, vegetation and soil) and artificial structures (buildings and roads). The subimage was classified into four classes, including grassland, forest, water and built-up areas. The generated subpixel map with an overall accuracy of 97.3% was used as reference.

Fig. 3(a) shows the original image is degraded into a coarse-resolution image by a factor of 10 to spatial resolution of 300 m [see Fig. 3(b)]. In the coarse spatial resolution image, subpixels with the same attributes are totaled to obtain the pixel-level proportions of each class according to the generated high-resolution land cover map. Compared with the soft classification method, pixel-level proportions calculated by the prior land cover map allow us to better focus on the subpixel mapping algorithm and test its mapping accuracy. Fig. 3(c)–(f) show the obtained pixel-level proportions of different land cover types, including forest, vegetation, built-up areas and water, which provide the input of the proposed enhanced pixel-swapping mapping algorithm. The proposed algorithm was performed in an iterative manner. Fig. 3(g) shows the land cover map at year 2000, which was used as prior temporal information for the proposed algorithm. Fig. 3(k) shows the final subpixel map after 40 iterations. For comparison, results of the bicubic interpolation (BI), PSA, SLCCM and HC methods are shown in Fig. 3(h)–(l), respectively.

Two measures including the overall accuracy (OA) and Kappa coefficient are used to assess the quality of the obtained subpixel map by comparing the result with the reference land cover map. To evaluate the performance of STPSA when both small and great changes in the land cover occur, the proposed algorithm is tested at change/no change level. Fig. 3(k) shows the final result of STPSA with a scale factor of 10: it achieves the OA and Kappa coefficient as 73.6% and 53.6%, respectively, when focused on the fast change area, where the changes of forest/grassland to built-up areas and forest to grassland occur, while OA and Kappa coefficients for the unchanged area are 94.9% and 92.2%, respectively. For comparison, the overall accuracy of the mapping results with the original PSA, BI, and HC methods without prior information in the changed area are also given: their OAs are 69.6%, 69.3%, and 67.4%, respectively.

Different zoom factors of 5, 10, and 15 are analyzed in this experiment, and the accuracy statistic for different approaches is shown in Table I, from which some significant findings are summarized below: 1) as the zoom factor increases, the overall accuracy of each approach decreases. The PSA in the unchanged area, for example, decreases in overall accuracy from 88.5% to 81.7%, then 76.6% as the factor increases from 5 to 10, then 15. This trend reflects the difficulty of this issue with a large factor; 2) comparing the performance of different approaches, the proposed STPSA is better than SLCCM, and then PSA, BI, and HC; 3) overall accuracies of changed areas are much lower than the unchanged area,
This is mainly due to the fact that changed areas tend to have a relatively high irregularity and heterogeneity, which increase the uncertainty for subpixel mapping; and 4) the HC method achieves preferable results for the unchanged area compared to its poor performance for the subpixel mapping task for changed areas. Accuracy analysis shows that it is better than PSA for the unchanged area, and the result is preferable to the result of SLCCM and STPSA when a small factor of 5 was used, which indicates that the HC method is a good choice when it is known that the area has no, or minor, land cover changes.

**IV. DISCUSSION**

**A. Strengths of STPSA**

As Fig. 4 displays the overall accuracy of different approaches for the changed area, it is apparent that the proposed algorithm STPSA performs better than the other methods: HC, BI, PSA, and SLCCM. Compared with other methods, both STPSA and SLCCM can obtain better results than them, especially with a large zoom factor. The superiority is due to the usage of prior temporal information. Compared with the performance of SLCCM and STPSA, STPSA performs slightly better than SLCCM, due to the fact that prior land cover proportional information used in SLCCM is restricted to a percentage level rather than the real subpixel level like STPSA. Correspondingly, the real spatial distribution for changed subpixels within each percentage cannot be predicted well in SLCCM, while ours improved it.

STPSA is also a well-designed algorithm to accommodate multi-type prior information like point-based sample data by field survey or image-based land cover maps. If accurate registration is performed, then the prior information can be inserted into the proposed algorithm by the defined neighborhood system and finally improves the subpixel mapping result.

**B. Choice of Parameter of STPSA**

The size of the neighborhood system is an important parameter and needed to be set by the investigator. A large neighborhood system tends to include more contextual information and improves the final result, but it may attract subpixels that are not neighbors in the real image, which may cause error. Meanwhile, a small neighborhood system can avoid this issue, but it may not take full advantage of contextual information. A study by Atkinson [3] showed that a larger number of neighbors tend to give a better result; the size 5 was recommended for the IKONOS image. It has not been confirmed whether it is a valid choice for the STPSA. In order to better understand and evaluate the effect of this parameter, the overall accuracy for STPSA and PSA with a different neighborhood size was computed.

Fig. 5(a) shows the accuracy statistic for PSA with a different neighborhood size, the accuracy tends to be stable and not sensitive when the neighborhood size is greater than four. When considering the neighborhood size, it should not be so large in order to avoid the attraction of irrelevant subpixels and preserve more spatial details. The neighborhood was set as six for the PSA algorithm in this study, and is also consistent with previous studies.

While the accuracy statistic for STPSA with a different number of neighbors is shown in Fig. 5(b), the overall accuracy decreases as the neighborhood size increases. An optimum result was obtained when the neighborhood size was set as two. The above test results also show that the optimum neighborhood size is different for STPSA and conventional PSA, and is due mainly to the fact that the influence of the spatial information is weakened by including temporal information, and its effective spatial distance was also shortened. The optimum neighborhood size for STPSA tends to be a tradeoff between temporal and spatial contextual information. Therefore, the optimum neighborhood size is different for PSA and STPSA. The neighborhood size for STPSA was set as two in this study.
C. Limitations of STPSA

Although the proposed algorithm has been well tested with actual satellite data, there are still some limitations. First, similar to the original PSA, this improved algorithm faces the problem of maximizing the spatial and temporal correlations between neighbors. Secondly, the weights of the surrounding pixels are the same in the proposed algorithm, but it is a simplified model of the actual situation. Actually, the weighting information of pixels at different times may vary, and the information at the same time at different locations may also be different. Thirdly, like the conventional PSA, in which spatial correlation information is utilized, STPSA is more suitable when the objects in the actual observations have high spatial correlation, e.g., most of the objects cover more than one pixel rather than many tiny targets within a pixel. Finally, high quality pixel-level proportions are requisite for high quality subpixel mapping with the proposed algorithm. A high quality soft classification model should be developed for high quality pixel mapping.

V. Conclusion

The STPSA was presented in this letter, in which spatial and temporal contextual information was applied to the subpixel mapping task in an area without high-resolution data. Experiments have demonstrated that the proposed algorithm performs better than HC, BI, PSA, and SLCCM algorithms. As prior temporal information is included in both STPSA and SLCCM, they perform better than other methods, especially with a large zoom factor. When compared with the performance of SLCCM and STPSA, the proposed STPSA performs slightly better than SLCCM, due to the fact that the subpixel allocating strategy used in SLCCM is restricted to a percentage level rather than the real subpixel level.

Adopting a method where swapping attractiveness is computed based on pairs of subpixels, rather than pixel by pixel as in the conventional algorithm, the proposed STPSA extends the applicability of conventional PSAs for binary land cover classes and multiple land cover classes. A specific spatial and temporal neighborhood system was constructed for the proposed algorithm to accommodate prior information such as previous land cover maps and field sampling land cover data, and the optimum neighborhood size was also discussed. Experiments showed that the optimum neighborhood size for STPSA is two while for conventional PSA it is six in this study where Landsat data was used.

As an important input of the proposed PSA, high quality land cover proportions are a prerequisite for obtaining a high quality subpixel map. Therefore, our further research will also include the development of more sophisticated approaches for providing high quality land cover proportions in order to improve subpixel mapping accuracy.

The authors would like to thank the two anonymous reviewers for their constructive comments which have helped improve this manuscript.

REFERENCES