

Improved Connected Component Labeling Algorithm for Remote Sensing Image Classification

M. Sumathi ^{*1}, T. Balaji ²

^{*1} Associate Professor, P.G. and Research Dept. of Computer Science, Sri Meenakshi Govt. Arts College for Women, Madurai, India

sumathivasagam@gmail.com

² Assistant Professor, Dept. of Computer Science, Govt. Arts College, Melur, India

bkmdgacm1976@gmail.com

ABSTRACT

Article Info

Volume 7, Issue 2

Page Number : 294-303

Publication Issue :

March-April-2021

Article History

Accepted : 12 April 2021

Published : 17 April 2021

The main objective of this paper is to carry out a detailed analysis of the most popular Connected Component Labeling (CCL) algorithms for remote sensing image classification. This algorithm searches line-by-line, top to bottom to assign a splotch label to each current pixel that is connected to a splotch. This paper presents two new strategies that can be used to greatly improve the speed of connected component labeling algorithms. It assigns a label to a new object, most labeling algorithms use a scanning step that examines some of its neighbors. The first strategy deeds the dependencies among the neighbors to reduce the number of neighbors examined. The second strategy uses an array to store the equivalence information among the labels. This replaces the pointer based deep rooted trees used to store the same equivalence information. It reduces the memory required and also produces consecutive final labels. The connected component labeling assigns labels to a pixel such that adjacent pixels of the same features are assigned the same label. The paper presents a modification of this algorithm that allows the resolution of merged labels and experimental results demonstrate that proposed method is much more efficient than conventional methods for various kinds of color images. This method is improving the labeling algorithms and also benefits for other applications in computer vision and pattern recognition

Keywords : CCL, 8DLA, I8DLA, Pixel Connectivity and Pixel Relationship

I. INTRODUCTION

The CCL (Connected Components Labeling) is an important step in many image processing applications.

There are typically four stages in such algorithms. First the input image is preprocessed through filtering and thresholding to segment the objects from the background. The preprocessed image is

usually binary, consisting of a number of regions against a background. Second, connected components labeling is used to assign each region a unique label, enabling the different objects to be distinguished. In the third stage, each region is processed to extract a number of features of the object represented by the region. In the final stage, these features are used to classify each region into one of two or more classes. The CCL is an important in the image classification processing. Two fundamental methods of connected components labeling are specified based on objects that handling: one processes image based on pixels while the other based on run-lengths. Both of them are facing the problem that much time is consumed just to eliminate labeling conflict.

In order to solve the labeling conflict, a relation table is adopted in traditional pixel-based connected components labeling method and more than two scans of image are needed. In the first scan, a temporary label number is assigned to each pixel belongs to any target. During this scanning, a relation table is created to records the equivalence relationship among these label numbers. Then at least another scan is needed to eliminate conflicting temporary label numbers in a connected component integrating with the previous relation table. This method is to implement but plenty of time is taken away in the several scans of image. Thus it is hardly suitable in the circumstances that have strict request for real-time performance. The improved connected components labeling method based on pixel by adding the rescan algorithm to eliminate the labeling errors stimulated by label conflicts. In their algorithm, only one scan is necessary to find and label all connected components but it is still time-consuming when lots of label conflicts are found or the connected components are high proportion of pixels. Other improved labeling algorithms devoted raise labeling speed but all of the methods are essentially

just to label every pixel. As a result, the label speed is serious affected by pixels of connected components.

Detection of connected components between pixels in binary images is a fundamental step in classification of an image objects and regions or splotches. Each splotch is assigned a unique label to separate it from other blobs. All the pixels within a blob of spatially connected 1's are assigned the same label. It can be used to establish boundaries of objects, components of regions and to count the number of blobs in an image. The original algorithm was developed by Rosenfeld and Pfaltz in 1966. It performs two passes through the image. In the first pass, the image is processed from left to right and top to bottom to generate labels for each pixel and all of the equivalent labels are stored in a pair of arrays. In the second pass, each label is replaced by the label assigned to its equivalence class. Several papers pointed out the problems in the second pass for large images because the equivalence arrays are resolved can have a dramatic effect upon the running time of this algorithm.

II. CONNECTED COMPONENT BASICS

In connected component-based region growing algorithm the input is grey scale and color images. The processing on different types of images consist of 3 main stages namely image pre-processing, selection using connected component and region growing that have been used in order to find out classification of satellite images.

A. Basic Pixel-Connectivity

It is used for establishing boundaries of objects and components of regions in image. Two pixels are said to be connected:

- if they are adjacent in some sense (neighbor pixels, 4/8/m-adjacency)

- if their gray levels satisfy a specified criterion of similarity(equal intensity level)

A pixel P at coordinate (x, y) has four direct neighbors $N_4(P)$ and four diagonal neighbors $N_D(P)$. The eight neighbors $N_8(P)$ of pixel P consist of the union of $N_4(P)$ and $N_D(P)$. There are three type of connectivity for pixels P and Q can be considered:

- Case 1: 4-connectivity – connected if Q is in $N_4(P)$
- Case 2: 8-connectivity – connected if Q is in $N_8(P)$
- Case 3: m-connectivity- connected if Q is in $N_4(P)$ or if Q is in $N_D(P)$ and $N_4(P) \cap N_4(Q) = \varphi$

Pixel connectivity is a central concept of both edge and region based approaches to classification. The notation of pixel connectivity describes a relation between two or more pixels. For two pixels to be connected they have to fulfill certain conditions on the pixel brightness and spatial adjacency.

First, in order for two pixels to be considered connected, their pixel values must both be from the same set of values V. Second, formulate the adjacency criterion for connectivity and introduce the notation of neighborhood. For a pixel P with the coordinates (x,y) the set of pixels given by:

$$N_4(P) = \{ (x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1) \}$$

$$N_8(P) = N_4(P) \cup \{ (x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1) \}$$

The two pixels P and Q, both having values from a set V are 4-connected if Q is from the set $N_4(P)$ and 8-connected if Q is from $N_8(P)$.

B. Pixel Relationship

A set of pixels in an image which are all connected to each other is called a connected component. Finding all connected components in an image and marking each of them with a distinctive label is called connected component labeling. It identifies the connected components in an image and assigning each one, a unique label, creating a Label Matrix Connected components labeling scans an image and

groups its pixels into components based on pixel connectivity, i.e. all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Since connected component labeling is a fundamental module in medical image processing, it improves the turn-around time of many medical diagnoses and procedures. An example of a binary image with two connected components which are based on 4-connectivity can be seen in Figure 1.

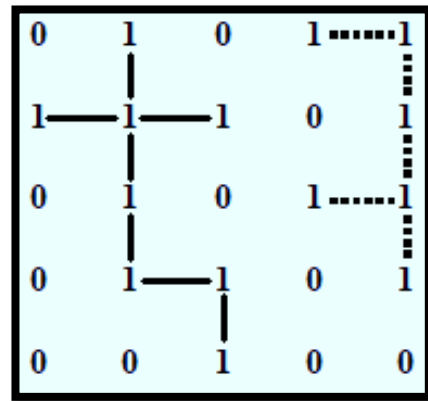
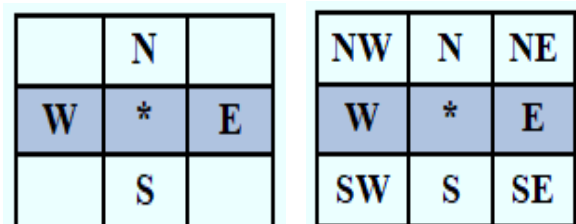


Figure 1. Four Connected Neighbor Points

The pixels of a binary image are 0's and 1's. The 1's will be used to denote foreground pixels and the 0's for background pixels. In image processing not only the value of a particular pixel is important, but also the values of its neighbors are used when processing that pixel. A pixel at coordinate $P(x, y)$ has four direct neighbors N_4P and four diagonal neighbors. The N_4P of $P(x, y)$ includes $P(x-1, y)$, $P(x+1, y)$, $P(x, y-1)$ and $P(x, y+1)$ referred as West, East, North and South (Fig. 1a) respectively. $N_D(P)$ of $P(x, y)$ includes $P(x-1, y-1)$, $P(x+1, y-1)$, $P(x-1, y+1)$ and $P(x+1, y+1)$ referred as Northwest, Northeast, Southwest, Southeast (Fig. 1b) respectively. The N_4P and N_DP are together known as 8-neighbors of the pixel P denoted by N_8P .



(a) (b)

Figure 2. a. Four Neighborhood Pixels, **b.** Eight Neighborhood Pixels

A pixel is a 4-neighbor of pixel $b(x, y)$ if it shares an edge with $b(x, y)$. The 4-neighbors of pixel $b(x, y)$, namely $P2, P4, P6$ and $P8$ are shown in Fig. 3. On the other hand, a pixel is an 8-neighbor of pixel $b(x, y)$ if it shares an edge or a vertex with $b(x, y)$. The 8-neighbors of pixel $b(x, y)$ namely $P1, P2, P3, P4, P5, P6$ to $P8$ are shown in Fig. 3. Usually 8-connectivity for connected component represents the foreground pixels and 4-connectivity for connected component hole represents the background pixels.

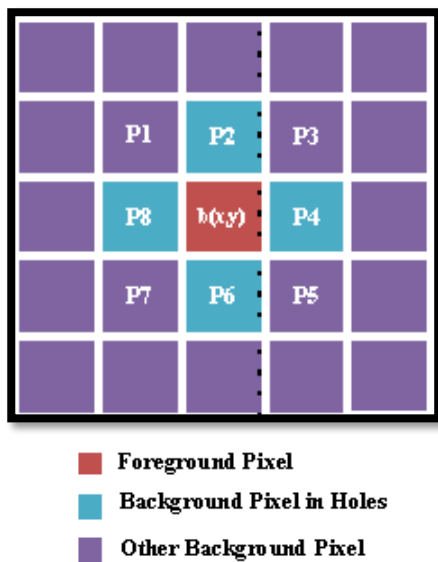


Figure 3. Neighboring of a Pixel Points

C. Simple CCL Algorithm

The connected component labeling works on gray level or color images and different measures of connectivity are possible. The connected components labeling operator scans the image by moving along a row until it comes to a point P (where P denotes the pixel to be labeled at any stage in the scanning process) for which $V=\{1\}$. When this is true, it examines the four neighbors of P which have already been encountered in the scan the neighbors. Based on this information, the labeling of P occurs as follows:

Case 1: If all four neighbors are 0 then assign a new label to point P

Case 2: If only one neighbor has $V=\{1\}$ then assign its label to P

Case 3: If one or more of the neighbors have $V=\{1\}$ then assign one of the labels to P and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class.

III. PROCESSING CHAIN OF CCL

The aim of this section is to perform connected component based classification with user defined classification criteria. The Processing chain can be divided in following steps:

- **Binary Mask Creation** - This mask is used as support of connected components classification (CC). The zero pixels are not taken into account by CC algorithm. Binarization criteria are defined by user, this step is optional, if no Mask is given, entire image will be processed.
- **Connected Component Classification** - Connected Component classification driven by user defined criteria.
- **Labelization and Object Characterization** - Connected Component Label Image is used to create label map, each Label object is characterized by shape and statistical measures.
- **Object Analysis Post Processing** - This step consists in post processing on each detected area of image labeling.
- **Output Vectorization** - Finally the results are exported in vector data shape format.

Many algorithms have been proposed for connected component labeling. They can be classified into three categories according to computer architecture and image representations.

A. Multi Scan Algorithms

In these algorithms, the image is scanned in the forward and backward raster directions alternately to propagate label equivalences until no label changes. This algorithm is designed for systems with limited

memory resources processing low resolution images. The performance of this algorithm is related to the size and the complexity of the binary image, thus it is not recommended for higher resolution images. Recent implementations improved processing time by introducing local equivalence tables, however the number of scans through the image frame is still dependent on the image complexity and is hard to predict. Hence these algorithms are not suitable for real-time video processing and also suitable for remote sensing image classification.

B. Two Scan Algorithms

The two scans are performed to complete the labeling process in certain criteria. In the first scan, provisional labels are assigned to object pixels, and the label equivalences are recorded. Label equivalences are resolved during or after the first scan. All equivalent labels are replaced by their representative label in the second scan. The improved version of two pass algorithm, very often referred to the classical algorithm, is used as a reference point in many benchmark tests. The key feature of this algorithm is the constant number of passes (two) through the binary image. Most of the two pass algorithms share similar features however they differ in data administration in image analysis.

C. One Scan Algorithms

These algorithms avoid analysis of label equivalences by tracing the contours of objects (connected components) or by the use of an iterative recursion operation. The most efficient single pass algorithm is relatively new group contour-tracing algorithm. They were developed specifically for labeling connected component in streaming data systems and classification analysis. The labeling step is performed in a single scan while data is streamed to the system. This ensures real-time processing speed. Also there is no need for buffering the input image frame; this results in lower memory requirements. The most significant feature of this algorithm can extract most

of the features of interest for all the objects during the scanning step, so there is no need to store a labeled image frame. This significantly reduces the memory requirements. This data is kept in a separate data array. Although results of the single pass labeling are sufficient for most object counting and pattern recognition systems. In order to label connected component in a single pass, the image has to be scanned in a raster scan processing.

IV. PROPOSED MULTI PASS TIME BASED CCL METHOD FOR CLASSIFICATION

In computer vision application, CCL is used for image classification to extract and label foreground pixels from background. Many approaches are proposed in the field of CCL. It presents two new multi-pass methods for matching CCL. They are 8DLA and I8DLA and are explained below in detail.

A. The 8DLA Method

This method utilizes the basic concept of 8-connectivity and checks all the 8-connected neighbors for minimum label. The virtual code for the 8-Directing Label Assortment method is given in Table I.

TABLE I

ALGORITHM FOR 8DLA METHOD

Procedure : 8DLA Method for RSIC

```

For j =1 to n iterations do
  For each pixel P in an image do
    If P is an object pixel
      Then it becomes a focused pixel
        For i = 1 to 8 directions do
          Search for minimum label
until
          background pixel is hit and put
it
          in miniature
        End for
      End if

```

```

Take minimum label m among miniature,
for 1<= i <= 8 and re-label matrix
    The focused pixel with the label is m
End for
If no label change for all the object pixels Then
exit
End for
    
```

Each pixel of the image is initialized with a unique label which is equal to the sequential index value of the pixel in the image. The iterations are the number of times the image is processed until each pixel is identified and labeled correctly. Every pixel in the input image is checked whether it is an object pixel having value 1. If so, then that pixel is considered as a focused pixel. Now keeping the focused pixel in mind all 8 directions around the focused pixel are checked one by one for minimum label. For example, we check the north direction first, and then the value of the pixel connected to the focused pixel in the concerned direction will be checked. If its value is 0 (meaning that it is a background pixel), then no further checking in that direction will be done. If the value of connected pixel is 1, then its label will be compared with current mini of the pixel in that direction. If the label is less than current mini then that label is stored in variable mini and the label of the next pixel in that direction is checked and so on until a background pixel having value 0 is hit. After that, remaining directions are checked with the same procedure.

After all 8-directions have been checked, a minimum label is obtained and then compared with the label of the focused pixel. The minimum two will be assigned as the label of the focused pixel. This process is repeated for all the object pixels in the image and stops when there is no label change for all object pixels.

B. Improved 8-Directing Label Assortment (I8DLA) Method

The basic algorithm for I8DLA is some difference that of 8DLA. The virtual code for the I8DLA algorithm is given in Table II.

TABLE II

Algorithm for Improved I8DLA Method

Procedure : I8DLA Method for RSIC

```

For i=1 to n iterations do
For each pixel P in an image do
    If P is an object pixel
        Then it becomes a focused pixel
            If (i >=2) and (label of P is not the
smallest)
                Then apply 8DLA
            End if
        End if
    End for
    If no label change for all the object pixels
        Then Exit
    End for
    
```

The only modification is that after second iteration the label of the focused pixel is checked if it is the smallest or not. If it is not, 8DLA is applied. Otherwise, no further processing is performed, thus saving the computation time. The inspection for smallest is done as follows. Let initial label image array be L (Label). Then after assigning initial label sequentially to an image L(i) becomes i, for i = 0 to T-1 (T means total number of pixels in the image). For focused pixel P having a label j, then check if L(j) is still j. If it is say that j is the smallest label or not. Otherwise it is not the smallest since it has already been changed, thus having a possibility of further change. If L(j) is changed to something else later, then apply 8DLA again to pixels having label j. Algorithm stops when there is no label change for all

the object pixels in the image. Thus this algorithm saves a great deal of computation time.

In I8DLA, smallest check is performed after two iterations. This number is chosen empirically. Many experiments can be conducted for various numbers of iterations and for various kinds of remote sensing test data and found that after two iterations most of object pixels already has smallest label they ought to have due to deep 8-directiing search characteristics in the method.

C. Flow Diagram

The proposed algorithm usually assumes some kind of local neighborhood to search for minimum label within that neighborhood and sometimes memorizes label equivalences. This is repeated over multiple iterations. To speed up the labeling procedure, there are two common strategies. One is to reduce the average number of times for checking the processed neighbor pixels in the first scan and the other is to resolve the label equivalences quickly by an efficient data structure.

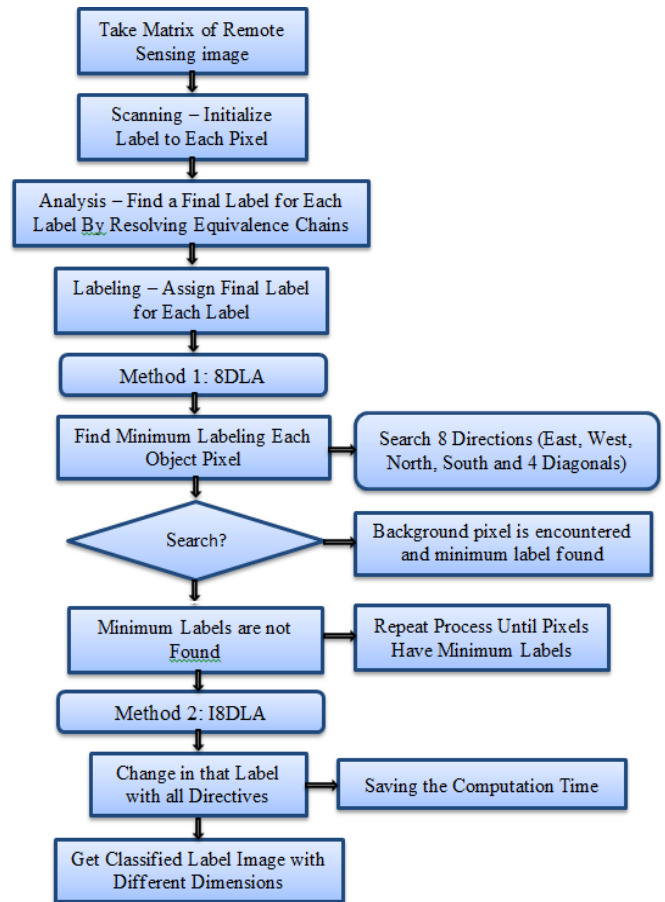
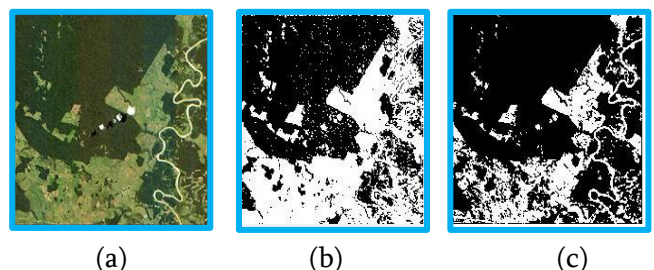


Figure 4. Flowchart for Proposed Multi Pass Time Based CCL Method

V. EXPERIMENTAL RESULTS

The experimentation of the algorithm is implemented in Mat lab 7.6 tool and is carried out on the remote sensing data set consisting of different images and as well as some images are taken from the web. Some of the images taken for the experiment are shown in Figure 5. The experimental results in this research confirm that the new proposed method labels the connected component and gives the number of components present in the binary images with 90 % and above accuracy.



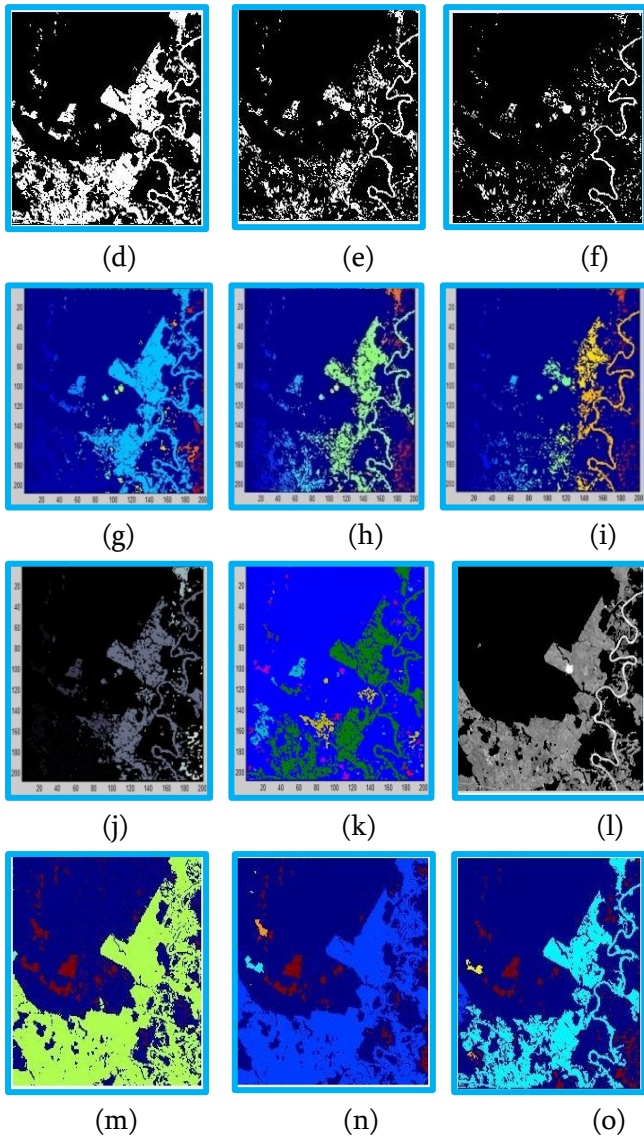


Figure 5. a. Original Image, b. BW Label Image 1, c. BW Label Image 2, d. BW Label Image 3, e. BW Label Image 4, f. BW Label Image 5, g. Label Selection Image 1, h. Label Selection Image 2, i. Label Selection Image 3, j. Label Selection Image 4, k. Label Selection Image 5, l. Classified Label Image 1, m. Classified Label Image 2, n. Classified Label Image 3, o. Classified Label Image 4.

VI. PERFORMANCE EVALUATION MATRICES

The results of the proposed multi pass time based CCL for RSI classification method is also compared with different class level of classified images from the remote sensing images and all the results have been reported in Table III.

TABLE III. PERFORMANCE EVALUATION MATRICES OF MULTI PASS TIME BASED CCL IN RSI CLASSIFICATION

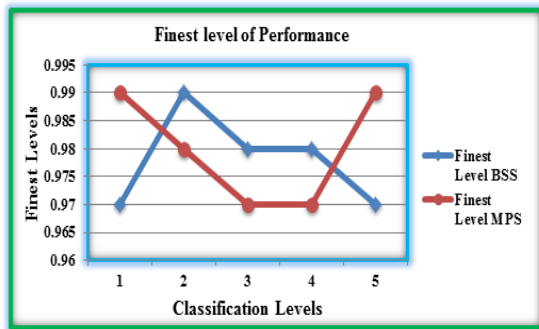
Output Image	Performance Matrices								
	PSNR	MSE	RMSE	NCC	NAE	MD	VIF	SDR	SC
Finest Level of Performance									
BWLabel Image 2	44.89	16.67	0.013	1.01	0.01	0	0.98	0.22	0.98
BWLabel Image 3	42.92	14.36	0.014	1.02	0.03	0	0.99	0.09	0.97
BWLabel Image 4	41.72	13.33	0.014	1.02	0.02	0	0.98	0.36	0.97
Label Selection Image 1	46.89	16.67	0.013	1.01	0.06	49	0.98	0.07	0.97
Label Selection Image 2	48.89	16.69	0.013	1.01	0.02	96	0.99	0.09	0.97
Label Selection Image 3	47.86	13.67	0.016	1.02	0.02	103	0.98	0.13	0.95
Label Selection Image 5	41.76	14.33	0.010	1.01	0.01	97	0.96	0.01	0.98
Classified Label Image 1	49.89	16.68	0.012	1.05	0.05	59	0.98	0.43	0.97
Classified Label Image 2	47.85	11.66	0.016	1.06	0.02	47	0.99	0.53	0.96
Classified Label Image 5	47.88	12.61	0.017	1.02	0.02	72	0.98	0.68	0.96
Average Level of Performance									
BWLabel Image 1	34.77	11.67	0.023	1.03	0.03	0	0.97	0.30	0.95
BWLabel Image 5	35.05	10.33	0.023	1.04	0.04	0	0.97	2.32	0.95
Label Selection Image 4	35.98	13.29	0.891	1.03	0.03	0	0.97	0.84	0.29
Classified Label Image 3	36.72	12.33	0.015	1.07	0.03	33	0.98	0.58	0.97
Classified Label Image 4	36.92	13.63	0.013	1.04	0.04	68	0.99	0.50	0.95

Output Image	Performance Matrices				Overall Efficiency	Computational Efficiency
	BSS	PBSS	MPS	PMPS		
Finest Level of Performance						
BWLabel Image 2	0.97	97.80 %	0.99	98.10 %	97.95 %	27.942517 Seconds
BWLabel Image 3	0.99		0.98			
BWLabel Image 4	0.98		0.97			
Label Selection Image 1	0.98		0.97			
Label Selection Image 2	0.97		0.99			
Label Selection Image 3	0.98	0.97				
Label Selection Image 5	0.99	0.98				
Classified Label Image 1	0.98	0.99				
Classified Label Image 2	0.97	0.99				
Classified Label Image 5	0.97	0.98				
Average Level of Performance						
BWLabel Image 1	0.92	93.00 %	0.93	93.20 %	93.10 %	
BWLabel Image 5	0.93		0.92			
Label Selection Image 4	0.92		0.94			
Classified Label Image 3	0.94		0.95			
Classified Label Image 4	0.94		0.92			

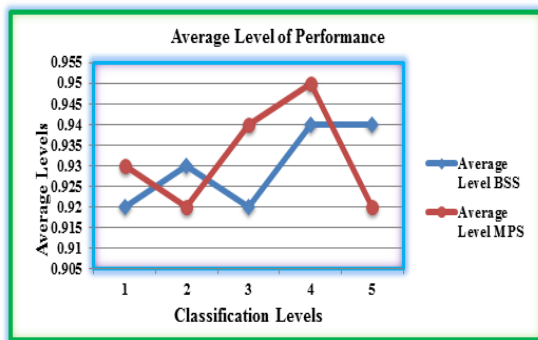
VII. COMPARISON GRAPH FOR TEST SCORES IN CCL

In order to improve the classification results, a nearest neighbor analysis by using class probabilities is performed for refining the class label of each connected component. The probabilities of connected

components within the selected regions are already computed during the classification. Already assigned class labels of the connected components are updated by using the hasty probabilities of connected components within selected region. The different types of metrics have been used for the performance evaluation of remote sensing image classification in improved CCL method. The CCL have been widely used in remote sensing vision applications. The multi pass time based improved CCL text score comparison graph is shown in the following Figure 6.



(a)



(b)

Figure 6. a. Graph for Finest Level of CCL, **b.** Graph for Average Level of CCL

VIII. CONCLUSION

A detailed analysis of these algorithms is provided in order to choose the most suitable for labeling and feature extraction from satellite images. The multi pass CCL algorithm proved to have the best capabilities. The performance of this algorithm is evaluated quantitatively by comparing the resulting of extractions with remote sensing images. The

results obtained show that the proposed method is able to identify in a robust way the I8DLA present in the satellite image, being able to properly classify all spatial features and offering a good level of overall accuracy.

The proposed multi pass time based CCL algorithm can work fast and accurately for most binary images and the optimizing of searching direction can improve its performance. Compared with other remote sensing classification algorithms by reference, the CCL algorithm makes better use of directional selection of labels and lengths have better performance in efficiency. The proposed technique is an important stage for the image directives method and found that the outcome of the algorithm is 96 %. Therefore it can say that our systematic method of classification gives meaningful, reasonable and effective classification for remote sensing images.

IX. DISCUSSIONS

The proposed multi pass time based CCL is an important step in remote sensing image classification and is often implemented in label selection framework to reduce execution time. It proposed two simple and fast multi pass CCL algorithms based on improving 8-directivesl label selection and showed that the proposed methods with various kinds of remote sensing test images. In order to confirm the algorithm, some simulations are conducted with this algorithm. The remote sensing testing images are shown in Fig. 6. Basic information of these test images performance matrices are shown in Table 3. By taking full advantage of the proposed algorithm also decreases time compared to that of algorithms proposed by other references. The reason for this phenomenon will be evident when we discuss the comparison results of the number of iterations required. The proposed labeling algorithm can also perform well and has a similar time consuming to

pixels based algorithm and are much better than that of previous reference.

X. REFERENCES

- [1]. Sutheebanjard P and Premchaiswadi W., "Efficient Scan Mask Techniques for Connected Components Labeling Algorithm", International Journal of Image Video Process, Vol. 44, pp. 1-20, 2018.
- [2]. He L., Chao Y and Suzuki K., "Fast Linear Time Two-Scan Labeling Algorithm", Proceedings of the IEEE International Conference on Image Processing, pp. 241-244, 2017
- [3]. Gayathri Devi G and Sumathi C. P., "Positional Connected Component Labeling Algorithm", Indian Journal of Science and Technology, Vol. 7, No. 3, pp. 306-311, 2014.
- [4]. Vyavahare A. J., "Connected Component Based Medical Image Classification", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 2, Issue 8, pp. 1-11, 2014.
- [5]. Li-Feng He and Yu-Yan Cha, "An Algorithm for Connected-Component Labeling, Hole Labeling and Euler Number Computing", Journal of Computer Science and Technology, Vol. 28, No. 3, pp. 468-478, 2013.
- [6]. Min Li, Xiaolin Zheng, "Classification of Brain Tissue based on Connected Component Labeling and Mathematic Morphology", IEEE Proceedings, pp. 978-989, 2011.
- [7]. He L, Chao Y and Suzuki K., "An Efficient First Scan Method for Label Equivalence based Labeling Algorithms", Pattern Recognition Letters, Vol. 31, No. 1, pp. 28-35, 2010.
- [8]. Wu K, Otoo E. and Suzuki K., "Optimizing Two Pass Connected Component Labeling Algorithms", Pattern Analysis & Applications, Vol. 12, No. 2, pp. 117-135, 2009.
- [9]. Roshan, Dharshana Yapa and Koichi Harada, "Connected Component Labeling Algorithms for Gray-Scale Images and Evaluation of Performance Using Digital Mammograms", International Journal of Computer Science and Network Security, Vol. 8, No.6, pp. 33-41, 2008.
- [10]. Suzuki K, Horiba I, and Sugie N., "Linear Time Connected Component Labeling based on Sequential Local Operations", Computer Vision and Image Understanding, Vol. 89, No. 1, pp. 1-23, 2007.

Cite this article as :

M. Sumathi, T. Balaji, "Improved Connected Component Labeling Algorithm for Remote Sensing Image Classification", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 2, pp. 294-303, March-April 2021. Available at doi : <https://doi.org/10.32628/CSEIT217262>

Journal URL : <https://ijsrcseit.com/CSEIT217262>