

Lựa chọn tham số t cho bài toán phân loại đám mây điểm LiDAR với thuật toán MCC

Nguyễn Thị Hữu Phương¹, Đặng Văn Đức², Nguyễn Trường Xuân¹

¹Khoa Công nghệ thông tin, Trường Đại học Mở - Địa chất

²Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

E-mail: ...

Tác giả liên hệ: Nguyễn Thị Hữu Phương, nguyenvhphuong85.nb@gmail.com

Ngày nhận bài: ..., ngày sửa chữa: ..., ngày duyệt đăng: ...

Online publication: ..., Digital Object Identifier:

Biên tập viên chính điều phối phản biện:

Tóm tắt: Thuật toán MCC được Evans và Hudak đề xuất vào năm 2007 để phân loại đám mây điểm LiDAR trong môi trường rừng. Hai tham số được sử dụng để khởi chạy MCC là tỷ lệ (s) và ngưỡng cong (t). Hai tham số này được xác định bởi người dùng và được chọn tùy thuộc vào mật độ của tia laser phản xạ trở lại. Trong nhiều nghiên cứu của các tác giả trên thế giới đã chỉ ra rằng việc lựa chọn thông số thích hợp t cho khu vực khảo sát sẽ ảnh hưởng đến độ chính xác của kết quả phân loại. Bài viết này giải quyết vấn đề chọn tham số t cho thuật toán MCC và thử nghiệm với đám mây điểm vùng Eel River.

Từ khóa: LiDAR; đám mây điểm; MCC; ngưỡng độ cong

Title: Select the parameter t for the LiDAR point cloud classification with MCC algorithm

Abstract: The MCC algorithm is an algorithm proposed by Evans and Hudak in 2007 to classify the LiDAR point cloud in a forest environment. The two parameters used to launch the MCC are the ratio (s) and the curvature threshold (t). These two parameters are defined by the user and are selected depending on the reflector laser density of the LiDAR system. In many studies of the world authors have shown that selecting the appropriate parameter t for the survey area will affect the accuracy of classification results. This paper addresses the issue of selecting the t parameter for the MCC algorithm and testing the Eel River area cloud.

Keywords: LiDAR, point cloud, MCC, curvature threshold.

I. GIỚI THIỆU

LiDAR là một công nghệ viễn thám mới, chủ động, có thể khảo sát trên mọi địa hình, đo được vào ban đêm và không phụ thuộc vào yếu tố thời tiết. Với khả năng khảo sát và đo trên một khu vực rộng lớn, hiện nay LiDAR đang được ứng dụng trong nhiều ngành và nhiều lĩnh vực của đời sống, xã hội như: trắc địa - bản đồ, khảo cổ học, vật lý,...

LiDAR là hệ thống bao gồm thiết bị laser, hệ thống GPS, và hệ thống INS. Hệ thống thiết bị laser được thiết kế phát các chùm tia laser, thu và nhận tín hiệu laser phản xạ, thu nhận dữ liệu cường độ tín hiệu laser phản xạ từ các đối tượng khác nhau trên mặt đất [1]. Từng tia laser riêng biệt không chỉ đi và về từ bề mặt Trái đất một cách trọn vẹn và duy nhất mà có thể được phản xạ thành một hay nhiều mức truyền. Mỗi mức phản xạ lại có một cường độ tín hiệu phản xạ khác nhau [1]. Một tia laser phát ra có thể thu từ 4 - 5 tín hiệu phản hồi. Tập hợp các điểm của các tia phản xạ này ta sẽ có đám mây điểm. Đám mây điểm là tập hợp của

các điểm có tọa độ (x, y, z) . Một đám mây điểm thường rất lớn, khoảng từ vài trăm nghìn đến vài triệu điểm. Và đây là tập hợp dữ liệu rất có ích thu thập được từ mặt đất. Tuy nhiên, để có thể sử dụng một cách hiệu quả và triệt để được những thông tin cần phải tiến hành phân loại điểm nhằm chia tách những điểm có cùng đặc trưng thành từng lớp riêng biệt như lớp mặt đất, lớp nhà, lớp thực vật thấp, lớp đường dây điện,...

Hiện nay, tại Việt Nam sử dụng công nghệ LiDAR mới chỉ dừng ở mức ứng dụng và sử dụng các phần mềm xử lý dữ liệu LiDAR được bán đi kèm thiết bị đo vẽ.

Trên thế giới, LiDAR được áp dụng từ khá sớm, có nhiều nhà khoa học đã nghiên cứu áp dụng và cải tiến thuật toán nhằm xử lý và thu nhận thông tin từ dữ liệu LiDAR một cách triệt để. Như trong các nghiên cứu của các tác giả [2-8].

Thuật toán MCC (Multiscale Curvature Classification) là thuật toán được đề xuất bởi Evans và Hudak vào năm

2007 để phân loại đám mây điểm LiDAR thành hai lớp là mặt đất và không mặt đất. Mục đích của thuật toán là lọc những tín hiệu phản xạ không mặt đất được phản hồi các đối tượng như thực vật, tòa nhà, đường dây điện,... từ đám mây điểm LiDAR để giữ lại những điểm thuộc lớp mặt đất phục vụ cho công tác thành lập DEM/DSM/DTM của khu vực khảo sát.

Trong nghiên cứu tác giả Evans và Hudak đã tiến hành phân loại tín hiệu phản xạ LiDAR trong môi trường rừng, thuật toán sử dụng phép lặp TPS (Thin Plate Spline) với lõi lọc 3×3 và kiểm soát khoảng cách các điểm mẫu ảnh hưởng đến bề mặt ước lượng [9]. Triển khai thuật toán MCC các tác giả đã định nghĩa một vectơ $Z(s)$ chứa giá trị tọa độ của các điểm LiDAR. Vector này sẽ được sử dụng trong phép lặp bề mặt sử dụng TPS với độ phân giải theo tỉ lệ λ . Dung sai độ cong t sẽ được thêm vào $x(s)$ và các điểm không mặt đất sẽ được loại bỏ khi $Z(s) > c$. λ và t là hai giá trị do người dùng tự định nghĩa.

Trong tài liệu [10] lại tiến hành so sánh thuật toán MCC và BCAL (Boise Center Aerospace Laboratory LiDAR) theo độ phân giải khác nhau của mô hình DTM cần thành lập. Các tham số khởi tạo được sử dụng để thành lập DTM là các giá trị sẽ gây ra lỗi RMSE cho nghiên cứu. MCC sẽ hoạt động tốt nhất với giá trị tỉ lệ là 1.0 và độ cong là 0.05.

II. PHÂN LOẠI Đám Mây Điểm LIDAR VỚI THUẬT TOÁN MCC

1. Dữ liệu đám mây điểm LiDAR

Công nghệ LiDAR sử dụng ánh sáng laser hồng ngoại để đo khoảng cách địa hình và các đối tượng. Xung laser được gửi đến mặt đất, bằng cách đo thời gian cho đến khi xung trở về, tính toán khoảng cách tới đối tượng. Bằng cách thu thập hàng ngàn điểm mỗi giây ta có mô hình mặt đất, cây cối, nhà cửa, đường giao thông và các đối tượng với độ chính xác cao và mật độ điểm.

liệu điểm LiDAR là dữ liệu 3 chiều (x, y, z) chứa các thông tin của bề mặt địa hình, bề mặt các công trình xây dựng, thực phủ, đường dây tải điện, đường giao thông,... Thuộc tính của điểm LiDAR như: cường độ, tọa độ (x, y, z) , góc quét, hướng quét, thời gian đo GPS,... Dữ liệu điểm được xử lý sau khi thu thập với công nghệ LiDAR thành điểm có tọa độ (x, y, z) chính xác cao bằng cách phân tích khoảng thời gian, góc tia quét, vị trí GPS và thông tin INS (đạo hàng quán tính - Inertial Navigation System).

Dữ liệu đám mây điểm thường rất lớn từ vài ngàn đến vài chục triệu điểm tùy thuộc vào độ rộng của khu vực đo vẽ và sự phức tạp của khu vực đo vẽ. Định dạng đám mây điểm thường là .las hay .laz. Định dạng .las là định dạng tệp chuẩn theo ngành được định nghĩa bởi Hiệp hội

ảnh hàng không và viễn thám Hoa kỳ (American Society of Photogrammetry and Remote Sensing). Một tập dữ liệu .las sau khi được xử lý có thể là các điểm thuộc lớp đất trống, thảm thực vật cao, thực vật thấp, nhà,...

2. Thuật toán MCC trong phân loại đám mây điểm

Đây là thuật toán phân loại tự động được sử dụng để phân loại đám mây điểm 3D LiDAR, phát triển bởi Moscow Forestry Science Laborator, sử dụng chủ yếu để phân loại dữ liệu LiDAR trong môi trường rừng. Kết quả của thuật toán là phân chia đám mây điểm thành hai lớp là mặt đất và không mặt đất [9].

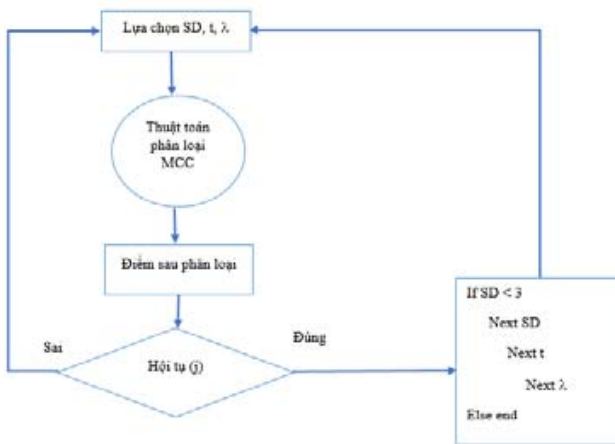
Nguyên tắc cơ bản của thuật toán là phân loại các điểm mà đi chệch khỏi một ngưỡng mà người dùng lựa chọn (t) được nội suy từ đám mây điểm với phép lặp TPS. Thuật toán sẽ lặp qua nhiều bước cho đến khi hội tụ.

Mô tả thuật toán [8]:

Các tham số khai báo:

- o SD - miền tỉ lệ, số nguyên trong khoảng [1..3]
- o t_{SD} : dung sai độ cong cho SD
- o t_1 do người dùng định nghĩa
- o $t_2 = t_1 + 0,1$
- o $t_3 = t_3 + 0,1$
- o CR_{SD} độ phân giải của miền tỉ lệ - λ
- o $\lambda_1 = 0,5 * \lambda$
- o λ_2 - người dùng định nghĩa (thường là khoảng cách sau của dữ liệu đầu vào LiDAR)
- o $\lambda_3 = 1,5 * \lambda$
- o P_U - tập điểm chưa phân loại, $P_U = \{P_1, P_2, \dots, P_n\}$, trong đó n thay đổi sau mỗi lần lặp
- o P_j - điểm LiDAR rời rạc, $P_j = \{x_j, y_j, z_j\}$
- o P_0 - tập hợp đám mây điểm LIDAR khởi tạo bởi người dùng

- 1) Lựa chọn λ , t , đám mây điểm cần phân loại, tập tin chứa kết quả phân loại
- 2) Loại bỏ những điểm có cùng tọa độ (x, y) trong U_0 - gán chúng vào lớp không mặt đất
- 3) for SD = 1 to 3
- 4) repeat
- 5) Lặp bề mặt raster mới sử dụng TPS - S
- 6) Lọc bề mặt qua lõi lọc 3×3 qua S - S'
- 7) Với mỗi $P_j \in P_U$
- 8) if $z_j > S'(x_j, y_j) + t$ then
- 9) phân lớp P_j vào lớp không mặt đất và loại bỏ P_j khỏi P_U
- 10) until số điểm bị loại bỏ $< 0,1\% * n$
- 11) Phân lớp những điểm còn lại trong P_U vào lớp mặt đất



Hình 1. Mô tả thuật toán MCC trong phân loại đám mây điểm LiDAR

Tính năng của thuật toán MCC trong phân loại đám mây điểm LiDAR hỗ trợ cho quá trình lặp để tìm kiếm bề mặt thật của Trái đất với độ phân giải tương ứng với mẫu.

3. Tham số t trong thuật toán MCC

Ngưỡng độ cong t là một trong hai tham số khởi tạo của thuật toán MCC. Ngưỡng độ cong thể hiện góc lớn nhất của độ cong cho phép cho một điểm. Nếu góc này lớn hơn, quá trình lọc sẽ dừng lại và gán điểm vào lớp không mặt đất. Giá trị khởi tạo của t do người dùng tự định nghĩa, thông thường giá trị mặc định của t là 0,3 (nếu khoảng cách được đo bằng m) và bằng 1 (nếu khoảng cách được đo bằng feet) [???]

Theo [11] cài đặt tối ưu cho các tham số của thuật toán MCC phụ thuộc vào độ phân giải của khu vực khảo sát và mật độ điểm của đám mây điểm, của địa hình cũng như của thực vật tại khu vực đó. Trong các lần lặp của thuật toán, ngưỡng khoan độ cong (t) là tham số có ảnh hưởng nhất đến kết quả, do đây là tham số được sử dụng để so sánh và gán lớp các điểm vào lớp mặt đất hay không mặt đất.

Theo [9], t với SD1 được định nghĩa bởi người dùng, khi tiến hành lặp với phép lặp TPS và lõi lọc 3×3 sẽ định nghĩa một vector mới $x(s)$ trùng với $Z(s)$ bao gồm tọa độ (x, y) và giá trị trung bình của bề mặt. Sau đó t sẽ được thêm vào $x(s)$ và điểm sẽ được phân loại thành lớp mặt đất hay không mặt đất bằng cách áp dụng điều kiện:

If $Z(s) > c$ then điểm được gán vào lớp không mặt đất, với $c = x(s) + t$.

Có thể thấy, điều kiện trên xảy ra khi giá trị của t vượt hoặc xấp xỉ với λ , điểm sẽ bị gán vào lớp không mặt đất, do vector $x(s)$ được lọc thông qua lõi lọc 3×3 và được

tính toán dựa trên giá trị trung bình của bề mặt. λ được lựa chọn dựa theo khoảng cách sau của dữ liệu đầu vào, trong đó khoảng cách sau được tính toán dựa trên mật độ xung phản xạ trên một m^2 .

$$\lambda = 1/\sqrt{\text{mật độ xung}/m^2} \quad (1)$$

Theo nghiên cứu của các tác giả trong [12], nếu khoảng cách sau của các điểm dữ liệu đầu vào là 1m, tham số tỉ lệ lựa chọn là 1, thì ngưỡng độ cong lựa chọn nên là 0,3; 0,4; hay 0,5.

Nếu giá trị t quá lớn được áp dụng cho những khu vực có thực vật thấp sẽ gây ra lỗi cho quá trình phân loại. Hay những khu vực có địa hình phức tạp, việc lựa chọn một ngưỡng độ cong kết hợp với tỉ lệ được sử dụng để nhằm đạt được kết quả phân loại chính xác cao hơn.

Theo [11], để lựa chọn ngưỡng độ cong phù hợp với khu vực khảo sát, tốt nhất nên bắt đầu với $t = 0.3$, sau đó tiến hành tăng thêm 0.1 cho đến khi đạt được kết quả phân loại tốt nhất.

Vì vậy để có thể lựa chọn tham số tốt nhất cho khu vực khảo sát, cần tiến hành thử nghiệm và tìm ra giá trị t cho thuật toán MCC.

III. THỬ NGHIỆM VÀ ĐÁNH GIÁ

1. Dữ liệu thử nghiệm

Dữ liệu LiDAR thử nghiệm là dữ liệu được đo bởi LiDAR hàng không (ALS - Aerial Laser Scanning). Dữ liệu thu thập thông tin về khu vực sông Eel của Mỹ, chùm bước sóng là 1064nm, độ cao của tia quét là 45m, góc phương vị tia quét là 315 độ. Dữ liệu được đo vào tháng 7 năm 2014, được tải về từ trang web <http://opentopo.sdsc.edu>, đây là trang web được cung cấp bởi National Center for Airborne Laser Mapping (NCALM). Đám mây điểm bao gồm 943.843 điểm có tọa độ (x, y, z) , mật độ điểm quét là 2,16 điểm/ m^2 . Tọa độ của khu vực thử nghiệm là:

$$X_{min} = 420893.5611; Y_{min} = 4467227.0495$$

$$X_{max} = 421466.8754; Y_{max} = 4467690.4093$$

2. Thử nghiệm phân loại đám mây điểm với MCC

1) Tiến hành chạy phân loại MCC với các tham số được lựa chọn $s = 3$ và $t = 0,3$:

Cấu trúc câu lệnh mcc-lidar lựa chọn tham số:

```
\WindowsFormsApp1\WindowsFormsApp1\bin
```

```
\Debug>mcc-lidar -s 3 -t 0.3
```

```
"E:\NCS\ThuNghiemMCC\points2.las"
```

points.txt - Notepad

File	Edit	Format	View	Help
420970.22	4467687.82	191.09		
420969.81	4467687.12	182.36		
420964.45	4467683.51	134.83		
420961.59	4467684.12	140.62		
420957.99	4467683.57	131.59		
420954.84	4467683.75	131.91		
420951.47	4467683.58	127.79		
420948.25	4467683.67	126.92		
420944.84	4467683.48	122.53		
420941.56	4467683.48	120.57		
420938.48	4467683.75	121.99		
420938.29	4467683.50	118.84		
420935.18	4467683.76	120.03		
420932.24	4467684.23	124.01		
420929.58	4467685.03	132.22		
420929.31	4467684.69	127.86		
420927.38	4467686.38	147.45		
420925.99	4467688.64	174.21		
420925.74	4467688.34	170.41		
420925.40	4467687.92	165.08		
420922.99	4467688.91	175.73		
420919.14	4467688.19	164.54		
420916.32	4467688.68	168.87		
420915.92	4467688.23	163.11		

Hình 2. Đám mây điểm trước phân loại



Hình 3. Mô hình DEM của khu vực khảo sát

"E:\NCS\ThuNghiemMCC\output-2017-07-10-9-46-47.las" Reading 934.843 points

from "E:\NCS\ThuNghiemMCC\points2.las"

Chạy phân loại với 3 miền tỉ lệ: SD1, SD2, và SD 3. Trong đó, SD1 với 9 lần lặp, SD2 với 5 lần lặp, SD3 với 6 lần lặp. Trong mỗi miền tỉ lệ, khi số điểm được phân loại vào nhóm điểm không mặt đất < 10% tổng số điểm phân loại, vòng lặp sẽ dừng lại và chuyển sang miền tỉ lệ tiếp theo. Bổ sung tiêu chí chọn số lần lặp 9, 5, 6. Kết quả phân loại:

output-2017-07-10-9-46-47.txt - Notepad

File	Edit	Format	View	Help
420996.45	4467690.38	140.30	1	
420999.63	4467690.34	141.80	1	
420999.05	4467689.45	143.17	1	
420995.92	4467689.65	143.73	1	
420995.78	4467689.31	139.49	2	
420992.96	4467690.19	148.41	1	
420986.40	4467689.90	140.90	1	
420983.20	4467689.92	139.12	1	
420980.17	4467690.31	142.00	1	
420949.91	4467690.39	128.33	1	
420981.94	4467689.65	139.18	1	
420985.15	4467689.61	140.73	1	
420991.82	4467690.16	151.53	1	
420994.67	4467689.40	143.98	1	
420997.77	4467689.15	142.86	1	
420998.24	4467688.43	142.29	1	
420995.16	4467688.72	143.84	1	
420995.00	4467688.34	139.17	1	
420992.20	4467689.25	148.57	1	
420991.95	4467688.70	141.66	1	
420988.81	4467688.86	141.75	1	
420985.64	4467688.97	141.08	1	
420985.51	4467688.69	137.58	1	

Hình 4. Các điểm được gán lớp sau phân loại với 1 - điểm không mặt đất, 2 - điểm mặt đất



Hình 5. Tỉ lệ phân loại trong các miền tỉ lệ

```
C:\Users\MyPC\My Documents>mcc-lidar
-s 3 -t 0.5 points2.las output.las
9 points (0.0055%) classified as non-ground
Classifying 164.227 points as ground...
Copying classifications to 3 duplicate points...
Writing output file "output.las"...
```

Hình 6. Kết quả số điểm mặt đất sau phân loại với t = 0.5

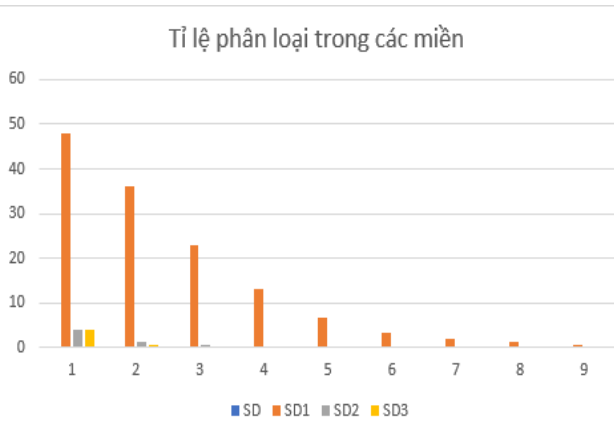
2) Tiến hành chạy phân loại với s = 3, t = 0,5:

Khởi tạo giá trị tham số cho s và t:

Chạy phân loại với 3 miền tỉ lệ: SD1 9 lần lặp, SD2 6 lần lặp, SD3 5 lần lặp. Tỉ lệ phân loại trong mỗi miền được thể hiện qua biểu đồ:

3) Với s = 3, t = 0,7:

Lựa chọn tham số cho phân loại: Chạy phân loại với t = 0,7 với 3 miền tỉ lệ: SD1 9 lần lặp, SD2 5 lần lặp, SD3



Hình 7. Tỉ lệ phân loại trong mỗi miền với $t = 0.5$

```

output1.txt - Notepad
File Edit Format View Help
420996.45 4467690.38 140.30 1
420999.63 4467690.34 141.80 1
420999.05 4467689.45 143.17 1
420995.92 4467689.65 143.73 1
420995.78 4467689.31 139.49 2
420992.96 4467690.19 148.41 1
420986.40 4467689.90 140.90 1
420983.20 4467689.92 139.12 1
420980.17 4467690.31 142.00 1
420949.91 4467690.39 128.33 1
420981.94 4467689.65 139.18 1
420985.15 4467689.61 140.73 1
420991.82 4467690.16 151.53 1
420994.67 4467689.40 143.98 1
420997.77 4467689.15 142.86 1
420998.24 4467688.43 142.29 1
420995.16 4467688.72 143.84 1
420995.00 4467688.34 139.17 2
420992.20 4467689.25 148.57 1
420991.95 4467688.70 141.66 1
420988.81 4467688.86 141.75 1
420985.64 4467688.97 141.08 1
420985.51 4467688.69 137.58 2
420950.37 4467689.73 128.73 1
420947.48 4467690.27 133.53 1
420944.05 4467690.03 128.52 1
    
```

Hình 8. Các điểm được gán sau phân loại

```

mcc-lidar -s 3 -t 0.7 points2.las
output.las
    
```

```

16 points (0.0091%) classified as non-ground
Classifying 175.873 points as ground..
Copying classifications to 3 duplicate points..
Writing output file ""output.las""...
    
```

Hình 9. Số lượng điểm thuộc lớp mặt đất

5 lần lặp. Tỉ lệ phân loại qua mỗi lần lặp của 3 miền tỉ lệ được thể hiện qua sơ đồ sau:

3. Đánh giá

Qua quá trình thử nghiệm với bộ dữ liệu 943.843 điểm, thử nghiệm 3 lần với giá trị tham số $s = 3$, giá trị $t = 0,3$; $t = 0,5$ và $t = 0,7$.

```

output2.txt - Notepad
File Edit Format View Help
420996.45 4467690.38 140.30 2
420999.63 4467690.34 141.80 1
420999.05 4467689.45 143.17 1
420995.92 4467689.65 143.73 1
420995.78 4467689.31 139.49 2
420992.96 4467690.19 148.41 1
420986.40 4467689.90 140.90 1
420983.20 4467689.92 139.12 1
420980.17 4467690.31 142.00 1
420949.91 4467690.39 128.33 1
420981.94 4467689.65 139.18 2
420985.15 4467689.61 140.73 1
420991.82 4467690.16 151.53 1
420994.67 4467689.40 143.98 1
420997.77 4467689.15 142.86 1
420998.24 4467688.43 142.29 1
420995.16 4467688.72 143.84 1
420995.00 4467688.34 139.17 2
420992.20 4467689.25 148.57 1
420991.95 4467688.70 141.66 1
420988.81 4467688.86 141.75 1
420985.64 4467688.97 141.08 1
420985.51 4467688.69 137.58 2
420950.37 4467689.73 128.73 1
420947.48 4467690.27 133.53 1
    
```

Hình 10. Các điểm sau phân loại với $t = 0.7$



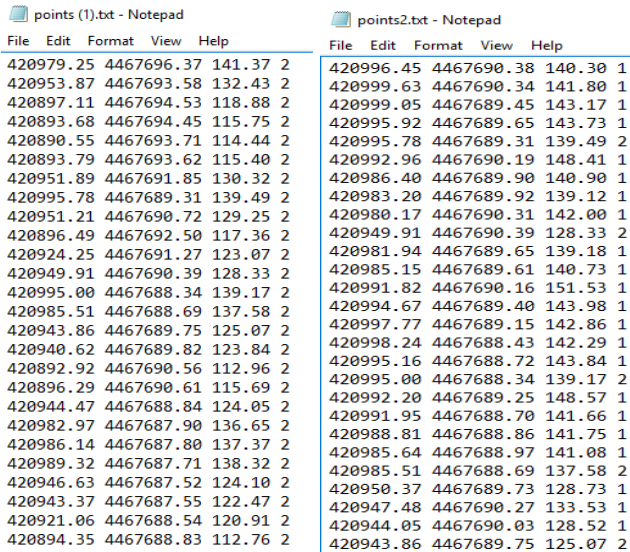
Hình 11. Tỉ lệ phân loại trong các miền với $t = 0.7$

Trong lần phân loại với $t = 0,3$ có 147.775 điểm được gán vào lớp mặt đất.

Phân loại với $t = 0,5$ có 164.227 điểm được gán vào lớp mặt đất.

Với $t = 0,7$ có 175.873 điểm được gán vào lớp mặt đất.

Để có thể đánh giá và lựa chọn được tham số t , so sánh kết quả sau phân loại với kết quả được chạy phân loại trang <http://opentopo.sdsc.edu>. Trên trang opentopo, đám mây điểm sau khi được thu thập, biên tập sẽ được phân loại với thuật toán BCAL (BCAL LiDAR tool). Đây là một mã nguồn mở được phát triển bởi Phòng thí nghiệm Boise Aerospace của Đại học Idaho. Đây là thuật toán phân loại dữ liệu LiDAR dựa trên thuộc tính độ cao, và thứ tự tia phản xạ, chủ yếu được sử dụng để phân loại lớp thực vật, như xác định mô hình tán cây. Tuy nhiên, thuật toán này sẽ cho kết quả phân loại không chính xác với những dữ liệu nằm ở gần biên.



Hình 12. Điểm thuộc lớp mặt đất được phân loại từ trang opentopo.sdsc.edu

Bảng I
SỰ THAY ĐỔI TỈ LỆ PHÂN LOẠI TRONG MỖI MIỀN
KHI CÓ SỰ THAY ĐỔI CỦA NGƯỠNG ĐỘ CONG

t	SD	Tỉ lệ phân loại								
t=0,3	SD1	49	37	24	13	7,4	4	2,2	1,3	0,79
	SD2	5,7	1,5	0,56	0,41	0,06				
	SD3	4,2	0,85	0,42	0,075	0,028	0,0041			
t=0,5	SD1	48	36	23	13	6,5	3,4	1,9	1,1	0,57
	SD2	4	1,3	0,55	0,31	0,26	0,03			
	SD3	3,9	0,53	0,2	0,0038	0,0055				
t=0,7	SD1	47	35	22	12	6,5	3,3	1,8	1,2	0,63
	SD2	3,4	1,1	0,47	0,12	0,034				
	SD3	3,3	0,55	0,085	0,028	0,0091				

Kết quả phân loại được chạy tự động với đám mây điểm trên trang <http://opentopo.sdsc.edu> có 166.280 điểm thuộc lớp mặt đất.

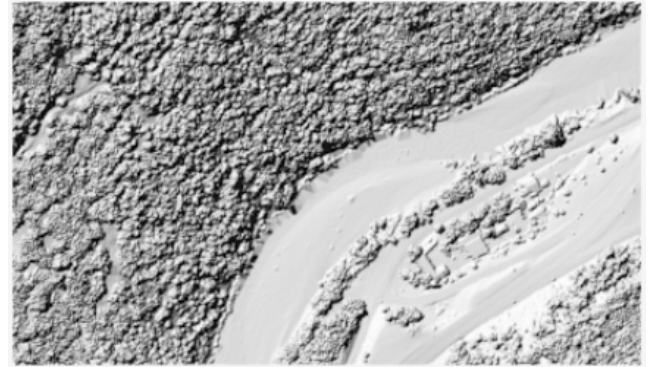
Số điểm lệch của từng kết quả thử nghiệm so với kết quả phân loại của Opentopo như sau:

- o t = 0,3 lệch 18.505 điểm, sai số 1,96
- o t = 0,5 lệch 2053 điểm, sai số 0,21
- o t = 0,7 lệch 9593 điểm, sai số 1,01

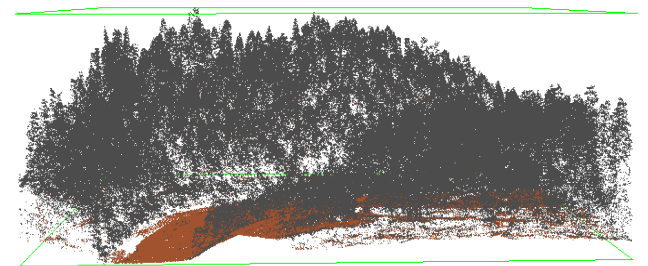
Sự thay đổi của ngưỡng độ cong còn thay đổi đến số lần lặp của mỗi miền tỉ lệ và điều kiện hội tụ của thuật toán cũng thay đổi với 3 giá trị t khác nhau.

Điều kiện hội tụ của SD1 là 0,0041 ; SD2 là 0,0055 và SD3 là 0,0091.

Trong 3 giá trị tham số t được lựa chọn cho s = 3 ta thấy với t = 0,5 kết quả phân loại có số điểm sai lệch là nhỏ nhất 2,1%.



Hình 13. Mô hình TIN của khu vực đo được thành lập sau phân loại đám mây điểm với MCC



Hình 14. Mô hình 3D của khu vực sông Eel

Từ đó, có thể thấy rằng tham số t sẽ quyết định đến quá trình phân loại một điểm thuộc lớp mặt đất hay thuộc lớp không mặt đất. Khi t thay đổi số lượng điểm được gán vào lớp mặt đất thay đổi. Và lựa chọn t phù hợp với miền tỉ lệ giúp cho quá trình phân loại có độ chính xác cao hơn.

Với dữ liệu thử nghiệm là khu vực ven sông El, với địa hình không phức tạp và nhiều thực vật, thuật toán MCC phân loại đám mây điểm LiDAR của khu vực đáp ứng được yêu cầu phân loại để thành lập DEM/DTM của khu vực.

IV. KẾT LUẬN

Bài báo đã trình bày thuật toán MCC trong phân loại đám mây điểm LiDAR, tiến hành thử nghiệm và lựa chọn giá trị tham số ngưỡng độ cong cho dữ liệu lựa chọn. Bộ dữ liệu với 943.843 điểm được phân loại với 3 giá trị tham số ngưỡng độ cong khác nhau. Sau khi phân loại, các điểm được gán vào lớp mặt đất (2) hay không mặt đất (1), nhận thấy rằng với s = 3 thì lựa chọn t = 0,5 là giá trị phù hợp để phân loại.

TÀI LIỆU THAM KHẢO

- [1] C. X. Tú, “Nghiên cứu ứng dụng công nghệ tích hợp LiDAR và máy ảnh số trong thành lập cơ sở dữ liệu mô hình số độ cao và bình đồ trực ảnh tại Việt Nam,” *Trường Đại học Mở - Địa chất, Hà Nội*, 2011.

- [2] K. Zhang, W. Bi, X. Zhang, X. Fu, L. Zhu, and K. Zhou, "A new kmeans clustering algorithm for point cloud," *International Journal of Hybrid Information Technology*, vol. 8, no. 9, pp. 157–170, 2015.
- [3] M. Rahman, B. Gorte, and A. Bucksch, "A new method for individual tree measurement from airborne lidar," in *Proc. of the SilviLaser*, 2009, pp. 283–288.
- [4] Z. Li, L. Zhang, X. Tong, B. Du, Y. Wang, L. Zhang, Z. Zhang, H. Liu, J. Mei, X. Xing *et al.*, "A three-step approach for tfs point cloud classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5412–5424, 2016.
- [5] H. Liu, L. J. Latecki, and W. Liu, "A unified curvature definition for regular, polygonal, and digital planar curves," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 104–124, 2008.
- [6] S. K. Lodha, D. M. Fitzpatrick, and D. P. Helmbold, "Aerial lidar data classification using expectation-maximization," in *Vision Geometry XV*, vol. 6499. International Society for Optics and Photonics, 2007, p. 64990L.
- [7] Y. Chang, A. Habib, D. Lee, and J. Yom, "Automatic classification of lidar data into ground and non-ground points," *International archives of Photogrammetry and Remote Sensing*, vol. 37, no. B4, pp. 463–468, 2008.
- [8] D. Gajski, T. Fiedler, and A. Krtalić, "Classification and filtering of airborne topographic lidar data," *The international archives of the photogrammetry, remote sensing and spatial information sciences (Print)*, vol. 34, no. 6/W11, p. 100, 2003.
- [9] J. S. Evans and A. T. Hudak, "A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 1029–1038, 2007.
- [10] W. T. Tinkham, H. Huang, A. Smith, R. Shrestha, M. J. Falkowski, A. T. Hudak, T. E. Link, N. F. Glenn, and D. G. Marks, "A comparison of two open source LiDAR surface classification algorithms," *Remote Sensing*, vol. 3, no. 3, pp. 638–649, 2011.
- [11] B. AHUDAK, "Miss."
- [12] BLUMENTRATH, "Miss."



Đặng Văn Đức sinh năm 1951 tại Hà Nội. Nhận học vị Tiến sĩ năm 1996. Nhận chức danh Phó giáo sư năm 2002.

Hiện công tác tại Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Lĩnh vực nghiên cứu : GIS và Viễn thám, Đa phương tiện, Công nghệ phần mềm.



Nguyễn Trường Xuân Sinh năm 1952 tại Hà Nội. Hiện công tác tại: Khoa Công nghệ thông tin, Trường Đại học Mỏ - Địa chất.

Hướng nghiên cứu: Hệ thống tin địa lý và Viễn thám, Địa tin học.



Nguyễn Thị Hữu Phương sinh năm 1985 tại Ninh Bình. Tốt nghiệp đại học tại Đại học Đại học Mỏ - Địa chất năm 2008. Tốt nghiệp Thạc sỹ tại Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội năm 2012.

Hiện đang công tác tại: Khoa Công nghệ thông tin, Trường Đại học Mỏ - Địa chất

Hướng nghiên cứu: Hệ thống thông tin, Cơ sở dữ liệu, Khai phá dữ liệu, Địa tin học.