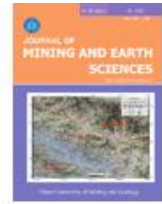




Journal of Mining and Earth Sciences

Website: <http://jmes.humg.edu.vn>



Building a program to automatically classify point cloud data



Quy Ngoc Bui ^{1,2,*}, Hien Dinh Le ³, Hiep Van Pham ^{1,2}, Tung Son Vu ⁴, Quan Anh Duong ^{1,2}, Trang Thu Thi Tran ¹

¹ Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, Vietnam

² Research and Development of Geospatial Data Management and Analysis Techniques (GMA), Hanoi University of Mining and Geology, Vietnam

³ Natural Resources and Environment One Member Co., Ltd, Hanoi, Vietnam

⁴ GeoPro Ltd, Hanoi, Vietnam

ARTICLE INFO

Article history:
Received 1st Apr. 2023
Revised 23rd July 2023
Accepted 17th Aug. 2023

Keywords:
3D city model,
LiDAR,
Point cloud,
Program.

ABSTRACT

Along with cartography science and technology development, data acquisition through aeronautical laser scanning systems has been developing. This is an essential and detailed data source for database construction, mapping, and city 3D modeling,... City 3D modeling requires processing many types of data, at which point cloud data processing and classification play an essential role in creating input data sources for the model. However, the processing and classification of point cloud data mainly depend on commercial software with very high costs; moreover, the algorithms and parameters of commercial software are locked. That makes it impossible for the user to intervene to improve product accuracy. Therefore, building a program to automatically classify point cloud data into different geographical objects helps us master data processing technology for creating 3D models. It makes an important contribution to building and developing smart cities. The article introduces the step-by-step classification of LiDAR point cloud data and the process of automatically building a program to classify point cloud data based on Visual Studio.Net language. The result is a bilingual program automatically classifying point cloud data (Vietnamese - English). The program can read and fully deploy algorithms to process LiDAR point cloud data containing information with four color bands (red, green, blue, and near-infrared). The primary processing is based on proposed classification steps and thresholds for point cloud data classification into eight feature classes, including hydrology, solar land, traffic, low plants, medium plants, high plants, houses, and other objects, to establish 3D city models.

Copyright © 2023 Hanoi University of Mining and Geology. All rights reserved.

**Corresponding author*

E - mail: buingocquy@humg.edu.vn

DOI: 10.46326/JMES.2023.64(4).01



Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <http://tapchi.humg.edu.vn>



Xây dựng chương trình tự động phân loại dữ liệu đám mây điểm

Bùi Ngọc Quý^{1,2,*}, Lê Đình Hiến³, Phạm Văn Hiệp^{1,2}, Vũ Sơn Tùng⁴, Dương Anh Quân^{1,2}, Trần Thị Thu Trang¹

¹ Khoa Trắc địa - Bản đồ và Quản lý đất đai, Trường Đại học Mỏ - Địa chất, Việt Nam

² Nhóm nghiên cứu nghiên cứu phát triển công nghệ quản lý và phân tích dữ liệu không gian địa lý, Hà Nội, Việt Nam

³ Công ty TNHH MTV Tài nguyên và Môi trường Việt Nam, Hà Nội, Việt Nam

⁴ Công ty TNHH GeoPro, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

TÓM TẮT

Quá trình:

Nhận bài 1/4/2023

Sửa xong 23/7/2023

Chấp nhận đăng 17/8/2023

Từ khóa:

Chương trình,
Đám mây điểm,
LiDAR,
Mô hình 3D thành phố.

Cùng với sự phát triển trong lĩnh vực khoa học và công nghệ đo đạc bản đồ, công tác thu nhận dữ liệu thông qua hệ thống quét laser hàng không đã có sự phát triển mạnh mẽ. Đây là nguồn dữ liệu quan trọng và chi tiết cho công tác xây dựng cơ sở dữ liệu, thành lập bản đồ, xây dựng mô hình 3D thành phố,... Quá trình xây dựng mô hình 3D thành phố đòi hỏi phải xử lý nhiều loại dữ liệu, trong đó công tác xử lý và phân loại dữ liệu đám mây điểm có vai trò quan trọng trong việc tạo nguồn dữ liệu đầu vào cho mô hình. Tuy nhiên, việc xử lý và phân loại dữ liệu đám mây điểm hiện nay chủ yếu phụ thuộc vào các phần mềm thương mại có bản quyền với giá thành rất cao, hơn nữa các thuật toán và tham số của phần mềm thương mại được mã hóa làm cho người dùng không thể can thiệp để cải thiện độ chính xác của sản phẩm. Do vậy, việc xây dựng một chương trình tự động phân loại dữ liệu đám mây điểm thành các đối tượng địa lý khác nhau giúp chúng ta có thể làm chủ được công nghệ xử lý dữ liệu phục vụ công tác xây dựng các mô hình 3D thành phố tiến tới góp phần quan trọng vào tiến trình xây dựng và phát triển thành phố thông minh. Bài báo giới thiệu quy trình các bước thực hiện phân loại dữ liệu đám mây điểm LiDAR và quá trình triển khai xây dựng chương trình tự động phân loại dữ liệu đám mây điểm dựa trên ngôn ngữ Visual Studio.Net. Kết quả bài báo đã xây dựng được chương trình tự động phân loại dữ liệu đám mây điểm song ngữ (Việt - Anh), có thể đọc và triển khai đầy đủ các thuật toán để xử lý dữ liệu đám mây điểm LiDAR chứa thông tin 4 kênh màu (đỏ, lục, lam và cận hồng ngoại). Quá trình xử lý chính dựa theo các bước và ngưỡng phân loại đã đề xuất cho phân loại dữ liệu đám mây điểm thành 8 lớp đối tượng địa lý khác nhau bao gồm: thủy văn, mặt đất, giao thông, thực vật thấp, thực vật trung bình, thực vật cao, nhà và các đối tượng khác phục vụ cho công tác thành lập các mô hình 3D thành phố.

© 2023 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

*Tác giả liên hệ

E - mail: buingocquy@humg.edu.vn

DOI: 10.46326/JM3ES.2023.64(4).01

1. Mở đầu

Hiện nay, trên thế giới các nghiên cứu về mô hình 3D thành phố phục vụ cho công tác mô hình hóa bề mặt và xây dựng thành phố thông minh đã và đang được chú trọng nghiên cứu và phát triển. Các mô hình 3D thành phố được xây dựng là kết quả của quá trình tích hợp nhiều loại dữ liệu khác nhau như: DEM, ảnh vệ tinh, bản đồ địa hình, ảnh số, dữ liệu đám mây điểm LIDAR,... (Bui và nnk., 2020; 2021; Dương và nnk., 2022; Lê, 2019; 2023). Quá trình xây dựng mô hình 3D thành phố đòi hỏi phải xử lý nhiều loại dữ liệu, trong đó công tác xử lý và phân loại dữ liệu đám mây điểm có vai trò quan trọng trong việc tạo nguồn dữ liệu đầu vào cho mô hình 3D thành phố. Nhiều công trình nghiên cứu đã đưa ra các phương pháp và thuật toán về phân loại đám mây điểm để từ đó nhận dạng các đối tượng bằng cách phân lớp cho đám mây điểm thu nhận được, trong đó chủ yếu tập trung vào 2 hướng: (1) Các thuật toán dựa trên mô hình máy học (Alexandre, 2020; Arief và nnk., 2019; Chenglu và nnk., 2019; Qiang và nnk., 2020; Yongguang và nnk., 2020); (2) Các thuật toán phân loại dựa trên hình thái và phổ (Heidar và nnk., 2020; Markus and Jing, 2014; Ronggang và nnk., 2018; Wuzhao và nnk., 2020; Yi và nnk., 2020; Yangbin và nnk., 2018; Xudong và nnk., 2019).

Thực tế hiện nay, công nghệ đo đạc thu thập dữ liệu đã dịch chuyển dần từ các thiết bị đo đạc truyền thống, đơn lẻ từng điểm sang các thiết bị thu thập thông tin không gian một cách toàn diện và nhanh chóng như máy quét laser mặt đất, máy quét laser di động (mobile mapping), thiết bị quét LiDAR trên máy bay hay trên UAV. Định dạng dữ liệu chung cho các thiết bị này là dữ liệu dạng đám mây điểm 3D mang thông tin chính xác về tọa độ địa lý và nhiều thông tin khác như màu sắc, cường độ phản xạ, xung phản hồi,... Với sự xuất hiện của dữ liệu đám mây điểm 3D, thế giới thực được thể hiện một cách đầy đủ trực quan với đúng tỷ lệ. Hơn nữa, khối lượng dữ liệu đám mây điểm 3D được thu thập ngày càng nhiều đã tạo điều kiện thuận lợi cung cấp nguồn thông tin đa dạng, đầy đủ cho phân loại và xây dựng các đối tượng nội dung của mô hình 3D thành phố phục vụ các ứng dụng quy hoạch, quản lý môi trường đô thị, không gian, cảnh quan,... Tuy nhiên, với khối lượng dữ liệu lớn, việc phân loại dữ liệu đám mây điểm hiện

nay hầu hết dựa trên các công cụ của phần mềm thương mại đi kèm của các hãng sản xuất thiết bị thu thập dữ liệu mà chưa có nhiều nghiên cứu xây dựng chương trình phân loại dữ liệu đám mây điểm. Hơn nữa, việc sử dụng các phần mềm thương mại đòi hỏi phải chi phí mua bản quyền lớn, các tham số và thuật toán tính toán được mã hóa làm cho người dùng không thể can thiệp để cải thiện độ chính xác của sản phẩm (Lê, 2023). Chính vì thế, việc thiết kế và xây dựng chương trình máy tính phục vụ công tác phân loại dữ liệu đám mây điểm có ý nghĩa khoa học và thực tiễn cao.

2. Thiết kế tổng thể chương trình phân loại dữ liệu đám mây điểm

2.1. Phương pháp phân loại dữ liệu đám mây điểm

Dữ liệu đám mây điểm LiDAR được thu nhận được với bốn dải màu: đỏ - xanh lục - xanh lam - hồng ngoại gần. Dựa trên các đặc điểm của đám mây điểm, nhóm nghiên cứu đã xây dựng quy trình chi tiết các bước để phân loại tự động dữ liệu đám mây điểm thành các lớp dữ liệu chứa các đối tượng địa lý khác nhau bao gồm: lớp điểm mặt đất, lớp điểm giao thông, lớp điểm thực vật, lớp điểm nhà và các lớp điểm các đối tượng khác. Đám mây điểm sau khi phân loại sẽ là dữ liệu đầu vào để thành lập mô hình 3D thành phố. Quy trình chi tiết bao gồm quá trình xử lý với các thuật toán lọc dữ liệu đám mây điểm thành các lớp đối tượng địa lý khác nhau. Việc phân loại dữ liệu đám mây điểm này dựa trên ngưỡng phân loại được mô tả trong Hình 1. Mức ngưỡng được phát triển dựa trên đặc điểm tự nhiên của các lớp đối tượng trong đám mây điểm cũng như cường độ phản xạ của các đối tượng địa lý được khuyến cáo từ hãng sản xuất thiết bị thu nhận dữ liệu (Bảng 1). Có bốn đặc điểm được sử dụng để phát triển mức ngưỡng đó là: cường độ, chỉ số thực vật (NDVI), chiều cao và hình học (Bui và nnk., 2021; Lê, 2023).

Bảng 1. Phạm vi cường độ của các đám mây điểm CityMapper cho các đối tượng khác nhau.

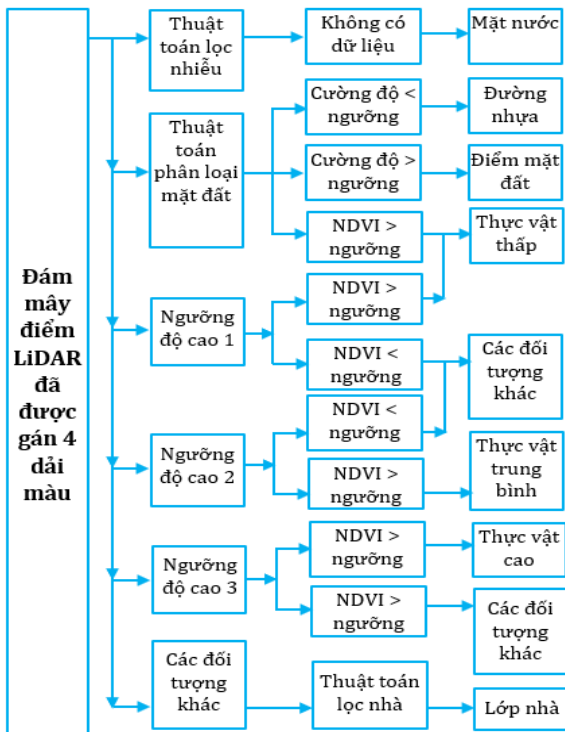
| TT | Đối tượng | Khoảng cường độ |
|----|------------|-----------------|
| 1 | Đường nhựa | 2100÷5400 |
| 2 | Bê tông | 9000÷12500 |
| 3 | Mái ngói | 7300÷9700 |
| 4 | Mặt đất | 7500÷13600 |

Bảng 1 cho thấy rằng đường nhựa có thể được lọc ra khỏi các vật thể mặt đất khác theo giá trị cường độ. Vì nền đất bằng phẳng trong khu vực gồm có đường nhựa, bê tông, ngói, đất ngô, đất.

Các bước triển khai phương pháp phân loại dữ liệu đám mây điểm cho xây dựng chương trình tự động phân loại dữ liệu đám mây điểm được quy trình hóa theo thứ tự triển khai thuật toán bao gồm: (1) Lọc bỏ điểm nhiễu; (2) Phân loại điểm mặt đất; (3) Phân loại giao thông; (4) Phân loại thực vật; (5) Phân loại điểm mái nhà (Dương và nnk., 2022).

2.2. Sơ đồ thiết kế hệ thống chương trình máy tính phân loại dữ liệu đám mây điểm

Chương trình phân loại dữ liệu được thiết kế để làm việc với dữ liệu đám mây điểm và dữ liệu số hóa (điểm, đường, vùng). Các định dạng dữ liệu này sẽ hiển thị trong môi trường 2D trên cửa sổ chính của chương trình với các công cụ hỗ trợ tương tác để người dùng có thể thay đổi các góc nhìn khác nhau của các đối tượng, biên tập và chỉnh sửa các dữ liệu trực tiếp trên chương trình đã xây dựng. Chương trình được thiết kế theo sơ



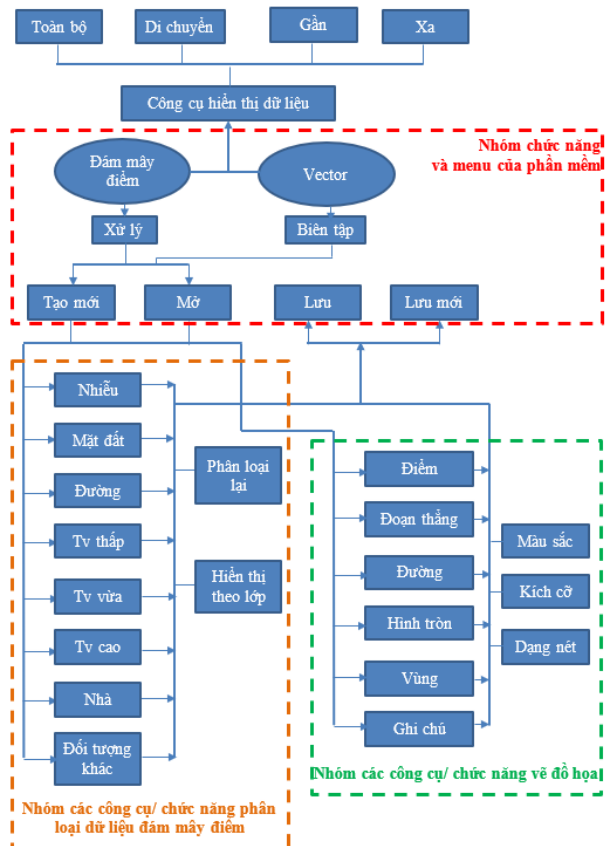
Hình 1. Quy trình phân loại các đối tượng từ dữ liệu đám mây điểm (Bui và nnk., 2021; Lê, 2023).

đồ thiết kế hệ thống với các nhóm chức năng và các công cụ như Hình 2, trong đó tập trung vào việc thiết kế nhóm chức năng phân loại dữ liệu đám mây điểm thành các lớp nội dung như đã đề cập trong mục 2.1.

3. Xây dựng chương trình phân loại dữ liệu đám mây điểm

3.1. Thiết kế giao diện chương trình

Chương trình được thiết kế trên nền tảng ngôn ngữ lập trình VisualBasic.Net với giao diện đơn giản dễ sử dụng, bao gồm 4 nhóm: (1) Menu bao gồm 6 mục chức năng chính: tệp, định dạng, vẽ, góc nhìn, công cụ và trợ giúp; (2) Các công cụ chức năng: tạo mới, mở tệp, lưu, các công cụ vẽ điểm, đường, hình tròn, ghi chú..., các công cụ thay đổi góc nhìn, các công cụ phân loại đám mây điểm, công cụ lọc, hiển thị và phân loại lại; (3) Màn hình hiển thị: cho phép vẽ các đối tượng đồ họa cơ bản như dạng điểm, đường, vùng và hiển thị dữ liệu



Hình 2. Sơ đồ thiết kế hệ thống chương trình phân loại đám mây điểm (Lê. 2023).

đám mây điểm trên giao diện của chương trình; (4) Thanh trạng thái phía dưới, hiển thị thông tin về tọa độ của con trỏ (Hình 3).

3.2. Thiết kế chức năng của chương trình phân loại dữ liệu đám mây điểm

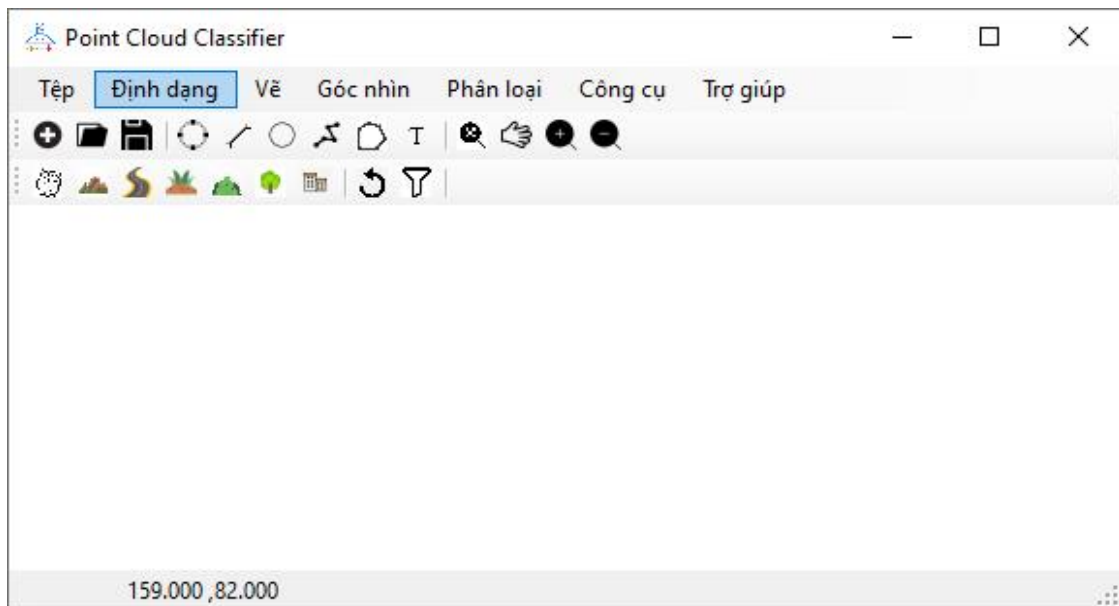
Chương trình được thiết kế với 7 nhóm chức năng bao gồm: (1) Nhóm chức năng quản lý tệp; (2) Nhóm chức năng hiển thị cho phép tùy chỉnh định dạng hiển thị của các đối tượng được số hóa bởi phần mềm; (3) Nhóm chức năng số hóa; (4) Nhóm chức năng điều chỉnh góc nhìn; (5) Nhóm các công cụ hỗ trợ như đo khoảng cách, công cụ đo diện tích, công cụ tra cứu thông tin đám mây điểm; (6) Nhóm công cụ trợ giúp giới thiệu về chương trình và 1 số thông tin hướng dẫn, đồng thời có thể lựa chọn chuyển đổi ngôn ngữ của chương trình qua lại giữa tiếng Việt và tiếng Anh; (7) Nhóm chức năng phân loại dữ liệu đám mây điểm bao gồm các công cụ thực hiện các tính năng phân loại đám mây điểm được thiết kế và xây dựng theo logic các bước đã đề xuất trong mục 2.1 bao gồm các công cụ: loại bỏ điểm nhiễu; phân loại điểm mặt đất; phân loại điểm đường giao thông; phân loại điểm thực vật; phân loại nhà; tái phân loại; lọc hiển thị theo lớp.

3.3. Lập trình xây dựng chương trình phân loại dữ liệu đám mây điểm

3.3.1. Khai báo cửa sổ hiển thị dữ liệu

Chương trình được thiết kế để phân định rõ giữa hệ tọa độ thực - r và hệ tọa độ màn hình - V. Hệ tọa độ thực lưu trữ thông tin tọa độ thực của mỗi đối tượng trong khi hệ tọa độ màn hình lưu trữ giá trị tọa độ hiển thị của các đối tượng trên cửa sổ hiển thị dữ liệu. Do đó, với bất kỳ đối tượng nào có trong dữ liệu đám mây điểm đầu vào, sẽ được lưu trữ thông tin dưới dạng cả 2 hệ tọa độ. Vì vậy, chương trình đã được xây dựng 6 hàm để chuyển đổi từ hệ tọa độ thực sang hệ tọa độ màn hình cho giá trị X,Y và hệ số tỷ lệ m.

```
'Cac tham so anh xa
Dim vXo, vYo As Integer
Dim rXo, rYo As Double
Dim m As Double
Dim rXmin, rYmin, rZmin As Double
Dim rXmax, rYmax, rZmax As Double
Dim rLasXmin, rLasYmin As Double
Dim rLasXmax, rLasYmax As Double
'Tu man hinh sang thuc
Private Function rX(ByVal vX As Double)
As Integer
    Return (rXo + (vX - vXo) / m)
End Function
Private Function rY(ByVal vY As Double)
As Integer
    Return (rYo - (vY - vYo) / m)
End Function
Private Function rS(ByVal vS As Double)
As Integer
    Return (vS / m)
```



Hình 3. Giao diện chương trình (Lê, 2023).


```

End Function
'Tu thuc sang man hinh
Private Function vX(ByVal rX As Double)
As Integer
Return (vXo + m * (rX - rXo))
End Function
Private Function vY(ByVal rY As Double)
As Integer
Return (vYo - m * (rY - rYo))
End Function
Private Function vS(ByVal rS As Double)
As Integer
Return (m * rS)
End Function

```

Trong 6 hàm trên, có rX, rY, rS, vS là các biến số, ứng với tọa độ thực và tọa độ màn hình của mỗi điểm; rXo, rYo là hằng số, mặc định ban đầu có giá trị là 0; vXo, vYo, m cũng là hằng số, được xác định khi mở cửa sổ chương trình. Mã nguồn lệnh mở cửa sổ chương trình và xác định các hằng số vXo, vYo, m được xác định như sau:

```

Private Sub FrmView_Load(ByVal sender As
Object, ByVal e As EventArgs) Handles
 MyBase.Load
g = PictureBox1.CreateGraphics
m = 1
vXo = PictureBox1.Width / 2
vYo = PictureBox1.Height / 2
End Sub

```

3.3.2. Khai báo khuôn dạng dữ liệu

Đối với lớp dữ liệu đường, polyline, hình tròn và ghi chú chương trình chỉ lưu trữ dưới dạng 2D do chương trình không thiết kế để hiển thị trong môi trường 3D. Với mỗi dạng đối tượng, tiến hành khai báo một mảng dữ liệu 2D, bao gồm các giá trị về tọa độ vị trí dưới dạng số thực, đồng thời với một giá trị số lượng dạng số nguyên tương ứng. Đường tròn khai báo thêm giá trị bán kính, còn ghi chú khai báo thêm giá trị nội dung ghi chú dưới dạng string.

```

Structure Line_2D
Dim rX1 As Double
Dim rY1 As Double
Dim rX2 As Double
Dim rY2 As Double
End Structure
Dim DS_Lines() As Line_2D, LinesCount As
Integer
Structure Circle_2D
Dim rX1 As Double
Dim rY1 As Double
Dim rR1 As Double
End Structure

```

```

Dim DS_Circles() As Circle_2D,
CirclesCount As Integer

```

Đối với định dạng dữ liệu polyline, tiến hành khai báo mảng polyline bao gồm mảng các vertex và số lượng vertex, cùng với số lượng polyline. Mỗi vertex sẽ chứa thông số về tọa độ X, Y từng vertex.

```

Structure Polyline_2D
Dim DS_Vertexs() As Point_3D,
VertexsCount As Integer
End Structure
Dim DS_Polylines() As Polyline_2D,
PolylinesCount As Integer

```

Ngoài ra, để phục vụ cho thuật toán phân loại điểm mặt đất, chương trình cần khai báo thêm mảng ô lưới (grid) và mảng tam giác (triangle). Mảng grid bên trong bao gồm các mảng con để lưu trữ các điểm trong mỗi ô lưới cùng với số lượng grid. Mảng triangle lưu trữ dưới dạng ba đỉnh của tam giác và số lượng tam giác.

```

Structure Grid_2D
Dim DS_Vertexs() As Point_3D,
VertexsCount As Integer
End Structure
Dim DS_Grids() As Grid_2D, GridsCount As
Integer
Structure Triangle
Dim id1 As Integer
Dim id2 As Integer
Dim id3 As Integer
End Structure
Dim DS_Triangles() As Triangle,
TrianglesCount As Integer

```

Đối với định dạng dữ liệu đám mây điểm, do dữ liệu LiDAR ngoài thông tin tọa độ X, Y, Z còn bao gồm thông tin về cường độ tia quét, lớp phân loại, thông tin về bốn dải màu: đỏ, lục, lam và cận hồng ngoại, nên cấu trúc định dạng đám mây điểm được khai báo theo mã nguồn lệnh khai báo định dạng dữ liệu cho đám mây điểm như sau:

```

Structure LasPoint_3D
Dim rX1 As Double
Dim rY1 As Double
Dim rZ1 As Double
Dim classification As Integer
Dim Intensity As Integer
Dim Red As Integer
Dim Green As Integer
Dim Blue As Integer
Dim NIR As Integer
End Structure

```

3.3.3. Khai báo đọc dữ liệu

Chương trình được xây dựng để đọc được dữ liệu đám mây điểm. Dữ liệu đám mây điểm cần được lưu trữ dưới dạng tệp văn bản, mỗi điểm tương ứng với một dòng, bao gồm các thông tin về tọa độ X, Y, Z, cường độ, thông tin màu sắc bao gồm giá trị đỏ, lục, lam, cận hồng ngoại và thông tin về phân lớp. Mỗi thông tin sẽ được ngăn cách bằng ký hiệu ngăn cách - delimiter thường là dấu cách trống hoặc dấu phẩy. Do đó, cần xây dựng một hàm giúp nhận diện ký hiệu ngăn cách này và lưu thông tin của từng điểm giữa mỗi ký hiệu ngăn cách. Đặt tên hàm này là SplitChain (tách chuỗi).

```
Function SplitChain(ByVal Text As String,
ByVal Delimiter As String, ByRef Words() As
String) As Integer
```

```
    Dim n As Integer, i As Integer
    i = 0
    Do
        i = i + 1
        Text = Trim(Text)
        n = InStr(Text, Delimiter)
        ReDim Preserve Words(i)
        If n > 0 Then
            Words(i) = Mid(Text, 1, n - 1)
            Text = Mid(Text, n + 1)
        Else
```

```
            Words(i) = Text
            Return i
        Exit Do
    End If
Loop
End Function
```

Nguyên lý hàm tách chuỗi này như sau: sử dụng lệnh lặp Do - Loop, với biến i chạy. Đầu tiên thực hiện cắt bỏ các khoảng trống thừa hai đầu mỗi dòng với hàm Trim(Text). Tiếp đến xác định giá trị số lượng chữ - n cho đến khi gặp ký tự ngăn cách - delimiter bằng hàm InStr. Với điều kiện n > 0, hàm sẽ lần lượt lưu các giá trị giữa các ký tự ngăn cách vào mảng Word(i) bằng hàm Mid(Text, 1, n-1). Như vậy các thông số về X, Y, Z, cường độ, thông tin màu sắc sẽ được lần lượt lưu vào mảng Word(i).

3.3.4. Khai báo hiển thị dữ liệu

Dữ liệu được hiển thị dưới dạng 2D, theo góc nhìn thẳng đứng từ trên xuống. Với mỗi đối tượng được vẽ, đều có lệnh Graphic.draw để vẽ các đối tượng với màu sắc mặc định là trắng trên nền đen. Với dữ liệu đám mây điểm *.las, sau khi được tải vào chương trình, dữ liệu sẽ được xử lý để tự động đưa toàn bộ dữ liệu về màn hình hiển thị (Hình 4).



Hình 4. Giao diện hiển thị dữ liệu đám mây điểm (Lê, 2023).

4. Kết quả và thảo luận

4.1. Kết quả xây dựng chương trình phân loại dữ liệu đám mây điểm

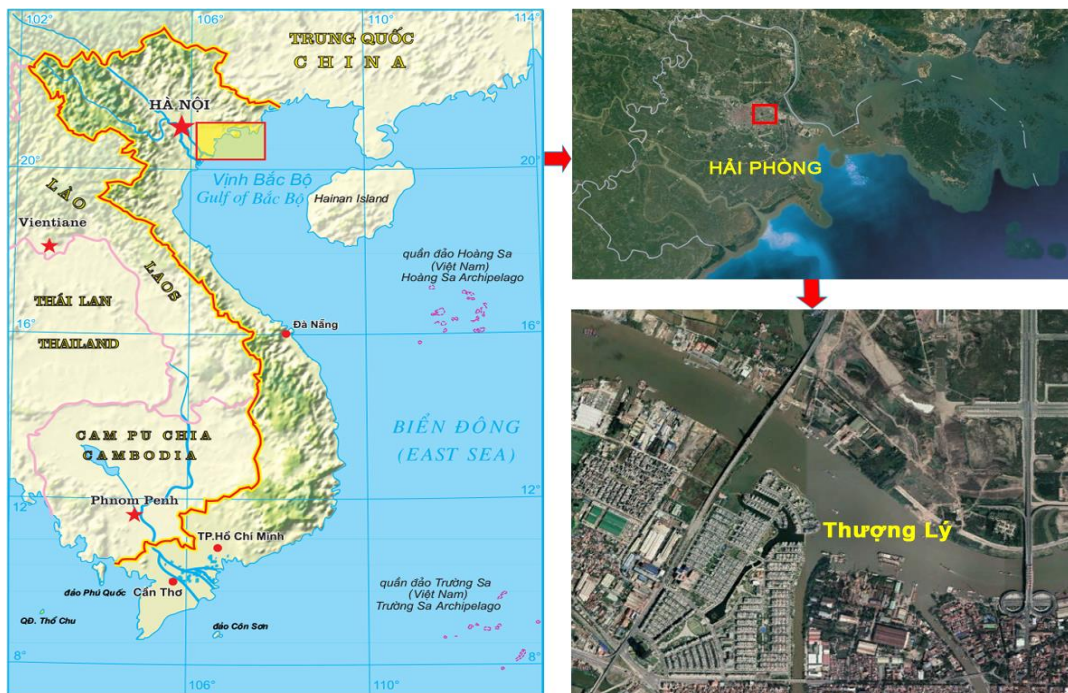
Với mỗi modul công cụ chức năng sau khi xây dựng sẽ được tiến hành chạy thử với dữ liệu đám mây điểm có số lượng từ nhỏ đến lớn để kiểm tra, nếu có phát sinh lỗi thì sẽ tiến hành kiểm tra lỗi và thực hiện các chỉnh sửa cần thiết. Mỗi modul công cụ sẽ được hoàn thiện sau khi chạy thử dữ liệu ở nhiều khu vực khác nhau và loại bỏ hết các trường hợp phát sinh lỗi. Các câu lệnh cũng được biên tập ra các chức năng riêng và bố trí một cách gọn gàng nhất, giúp tăng cường tốc độ xử lý dữ liệu, tránh các bước chạy đường vòng, lãng phí thời gian và tài nguyên bộ nhớ. Kết quả xây dựng chương trình có thể hiển thị và xử lý phân loại dữ liệu đám mây điểm để tách dữ liệu thành 8 lớp dữ liệu địa lý bao gồm: mặt nước, mặt đất, giao thông, thực vật thấp, thực vật trung bình, thực vật cao, nhà cửa và các đối tượng khác.

4.2. Ứng dụng chương trình trong phân loại dữ liệu đám mây điểm

Trên cơ sở chương trình đã xây dựng nhóm nghiên cứu tiến hành thực nghiệm phân loại dữ liệu đám mây điểm cho khu vực phường Thượng

Lý, thành phố Hải Phòng (Hình 5). Đám mây điểm được sử dụng công cụ lọc nhiễu để lọc bỏ các điểm thấp, kết quả thu được những khu vực không có dữ liệu và đưa về lớp thủy văn. Tiếp theo, tiến hành lựa chọn công cụ lọc mặt đất và thiết lập các tham số cho thuật toán lọc mặt đất bao gồm kích thước của ô lưới; khoảng cách và góc lặp lần 1; độ biến thiên của dữ liệu; khoảng cách và góc lặp lần 2 (Hình 6). Khi đó, chương trình sẽ căn cứ vào kích thước của ô lưới được lựa chọn để xác định các điểm hạt giống là các điểm thấp nhất trong các ô lưới và tiến hành tam giác hóa bề mặt để tạo mô hình TIN. Dựa trên bề mặt TIN này chương trình tiến hành áp dụng các điều kiện về khoảng cách và góc lặp để xác định điểm mặt đất, cụ thể điểm được xét là điểm nằm trong tam giác đồng thời thỏa mãn điều kiện nhỏ hơn giá trị tham số của khoảng cách và góc lặp mà người dùng đã thiết lập.

Khi đó, điểm này sẽ được tính là điểm mặt đất, chương trình sẽ xóa bỏ tam giác chứa nó và sử dụng điểm này để lập 3 tam giác giữa điểm đó và 3 cạnh của tam giác ban đầu. Sau khi vòng lặp 1 kết thúc chương trình sẽ tiến hành tương tự cho vòng lặp thứ 2 với các tham số đã được người dùng xác định với điểm khởi tính ban đầu là các điểm mặt đất của vòng lặp thứ nhất với giá trị góc lặp lớn hơn và khoảng cách lặp nhỏ hơn lần 1 để xác định các điểm mặt đất ở những khu vực địa hình có độ

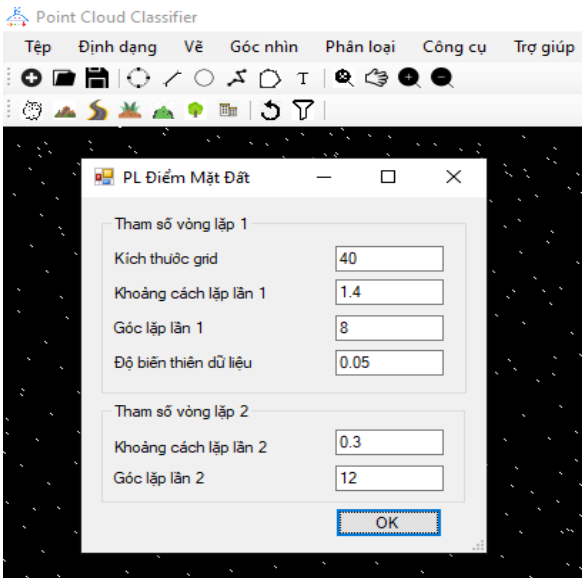


Hình 5. Vị trí khu vực phường Thượng Lý, thành phố Hải Phòng.

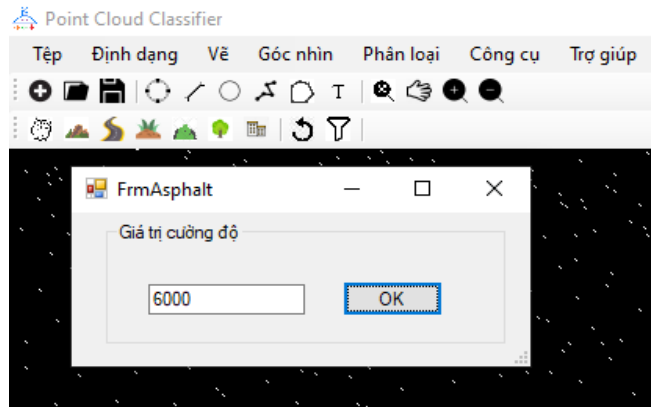
dốc lớn cũng như tránh cho bước nhảy lớn vào bề mặt của các đối tượng khác (Dương và nnk., 2022). Kết quả cuối cùng sẽ tách được lớp điểm mặt đất ra khỏi dữ liệu đám mây điểm gốc ban đầu (Hình 7). Sau khi phân loại xong các điểm mặt đất, tiến hành phân loại lớp đường giao thông dựa trên dữ liệu các điểm mặt đất đã được tách ra từ quá trình lọc mặt đất. Chương trình sẽ dựa trên tham số về giá trị cường độ mà người dùng nhập vào (Hình 8) để lọc ra các điểm đường giao thông. Kết quả của quá trình này sẽ tách được lớp mặt đất và lớp đường giao thông (Hình 9).

Sau khi lọc các đối tượng giao thông, tiến hành lọc các đối tượng thực vật. Thực vật được lọc ra thành 3 lớp (thấp, trung bình và cao) dựa trên giá trị NDVI kết hợp với độ cao của điểm so với bề mặt TIN được tạo ra từ kết quả phân loại điểm mặt đất. Kết quả của quá trình này sẽ thu được 4 lớp

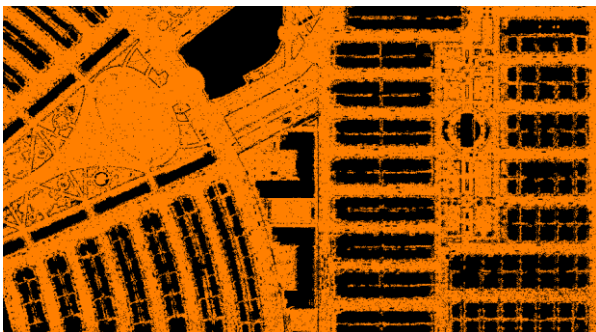
thực vật thấp, thực vật trung bình, thực vật cao và các đối tượng khác. Cuối cùng chương trình tiến hành triển khai thuật toán để phân loại lớp nhà trên cơ sở dữ liệu các đối tượng khác vừa được tách ra từ quá trình phân loại thực vật dựa vào đặc điểm dữ liệu lúc này đó là các điểm nằm trên một mặt phẳng đảm bảo thỏa mãn góc dốc so với bề mặt đất một giá trị được xác định của người dùng sẽ được đưa về lớp nhà. Kết quả sẽ thu nhận được lớp nhà (Hình 10) và các đối tượng địa lý còn lại sẽ được đưa vào lớp các đối tượng khác.



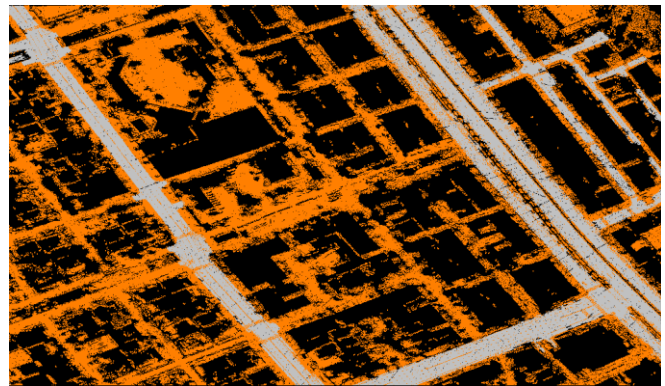
Hình 6. Xác định các tham số cho công cụ lọc điểm mặt đất.



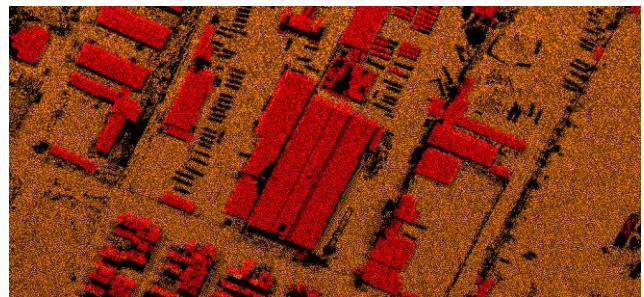
Hình 8. Xác định tham số cho lọc đường giao thông.



Hình 7. Kết quả phân loại lớp mặt đất (Dương và nnk., 2022).



Hình 9. Lớp giao thông được tách ra từ lớp điểm mặt đất dựa trên giá trị cường độ (Dương và nnk., 2022).



Hình 10. Kết quả phân loại lớp điểm mái nhà (Dương và nnk., 2022).

5. Kết luận

Bài báo đã tiến hành xây dựng và hiện thực hóa các bước triển khai phân loại dữ liệu đám mây điểm trên cơ sở thuật toán và ngưỡng phân loại được đề xuất bởi nhóm tác giả trong nghiên cứu của (Bùi và nnk., 2021; Lê, 2023). Kết quả xây dựng chương trình phân loại dữ liệu đám mây điểm của bài báo cho thấy chúng ta hoàn toàn có thể chủ động công tác thiết kế và xây dựng các chương trình máy tính hỗ trợ cho phân loại dữ liệu đám mây điểm độc lập mà không quá phụ thuộc vào các phần mềm thương mại như hiện nay.

Chương trình phân loại dữ liệu đám mây điểm đã xây dựng đảm bảo đọc được dữ liệu đám mây điểm có chứa thông tin 4 kênh màu bao gồm đỏ, lục, lam và cận hồng ngoại và có thể triển khai đầy đủ các thuật toán phân loại dữ liệu đám mây điểm theo quy trình các bước đã đề xuất và có thể phân loại dữ liệu đám mây điểm thành các đối tượng địa lý riêng biệt bao gồm: thủy văn, mặt đất, giao thông, thực vật thấp, thực vật trung bình, thực vật cao, nhà và các đối tượng khác. Chương trình có giao diện đơn giản, dễ nhìn, dễ sử dụng với 2 ngôn ngữ tiếng Việt và tiếng Anh. Đặc biệt, chương trình có thể lựa chọn các tham số cho từng đối tượng trong quá trình phân loại để phù hợp với từng nhóm đối tượng địa lý trong những điều kiện địa lý khác nhau giúp cho kết quả phân loại đảm bảo sát với thực tế nhất.

Với chương trình máy tính đã xây dựng, người dùng có thể đọc, hiển thị và tiến hành phân loại đám mây điểm ra nhiều các phân lớp khác nhau một cách tự động nhằm cung cấp dữ liệu đầu vào cho xây dựng các mô hình 3D thành phố. Ngoài ra, chương trình cũng cho phép số hóa các đối tượng đồ họa dạng điểm, đường, vùng hỗ trợ công tác đánh dấu các đối tượng khi cần thiết, đây cũng là một sự kết hợp cần thiết mà ít các phần mềm thương mại hiện nay quan tâm xây dựng.

Đóng góp của tác giả

Bùi Ngọc Quý - lên ý tưởng thiết kế, viết phần phương pháp và hoàn thiện bản thảo; Lê Đình Hiển - lập trình xây dựng chương trình; Dương Anh Quân - đánh giá các kết quả thực nghiệm và biên tập bản thảo; Phạm Văn Hiệp - chạy thử nghiệm và kiểm tra dữ liệu sau phân loại; Vũ Sơn Tùng - lọc dữ liệu, bắt lỗi chương trình; Trần Thị

Thu Trang - thu thập dữ liệu đám mây điểm thực nghiệm, biên tập và rà soát bản thảo.

Tài liệu tham khảo

- Alexandre Boulch (2020). ConvPoint: Continuous Convolutions for Point Cloud Processing. *Computers and Graphics*, 88, 24-34. doi.org/10.1016/j.cag.2020.02.005
- Arief, H. A. A., Indahl, U. G., Strand, G. H., & Tveite, H. (2019). Addressing overfitting on point cloud classification using Atrous XCRF. *ISPRS Journal of Photogrammetry and Remote Sensing*, 155, 90-101. doi.org/10.1016/j.isprsjprs.2019.07.002.
- Bui, N. Q., Le, D. H., Duong, A. Q., Nguyen Q. L. (2021). Rule-based classification of Airborne Laser Scanner data for automatic extraction of 3D objects in the urban area. *Journal of the Polish Mineral Engineering Society*, 48(2), 103-114. DOI: doi.org/10.29227/IM-2021-02-09
- Bui, N. Q., Le, D. H., Nguyen, Q. L., Tong, S. S., Duong, A. Q., Pham, V. H., Phan, T. H., Pham, T. L. (2020). Method of defining the parameters for UAV point cloud classification algorithm. *Journal of the Polish Mineral Engineering Society*, 46(1) 49-56. DOI: doi.org/10.29227/IM-2020-02-08.
- Chenglu Wen, Xiaotian Sun, Jonathan Li, Cheng Wang, Yan Guo, Ayman Habib (2019). A Deep Learning Framework for Road Marking Extraction, Classification and Completion from Mobile Laser Scanning Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 178-92. doi.org/10.1016/j.isprsjprs.2018.10.007
- Dương, A. Q., Lê, Đ. H., Phạm, V. H., Nguyễn, Q. C., Bùi, N. Q. (2022). Xây dựng quy trình thu nhận, xử lý và phân loại dữ liệu đám mây điểm LiDAR phục vụ thành lập mô hình 3D thành phố. *Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất*, 63(4), 1-12. [doi:10.46326/JMES.2022.63\(4\).01](https://doi.org/10.46326/JMES.2022.63(4).01)
- Heidar Rastiveis, Alireza Shams, Wayne A. Sarasua, Jonathan Li (2020). Automated Extraction of Lane Markings from Mobile LiDAR Point Clouds Based on Fuzzy Inference. *ISPRS Journal of Photogrammetry and Remote*

- Sensing*, 160, 149-66. doi.org/10.1016/j.isprsjprs.2019.12.009
- Lê, Đ. H. (2019). Nghiên cứu quy trình xử lý dữ liệu thu nhận từ hệ thống bay chụp ảnh & quét Lidar Leica City Mapper trong thành lập mô hình Cyber City. Luận văn Thạc sĩ Kỹ thuật, Trường Đại học Mỏ - Địa chất. Hà Nội, 82 trang. (Việt Nam).
- Lê, Đ. H. (2023). Nghiên cứu tối ưu hóa thuật toán tự động phân loại dữ liệu đám mây điểm hỗ trợ xây dựng mô hình 3D thành phố thông minh. Luận án Tiến sĩ Kỹ thuật, Trường Đại học Mỏ - Địa chất. Hà Nội, 135 trang. (Việt Nam).
- Markus Gerke, Jing Xiao (2014). Fusion of Airborne Laserscanning Point Clouds and Images for Supervised and Unsupervised Scene Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 78-92. doi.org/10.1016/j.isprsjprs.2013.10.011.
- Qiang Lu, Chao Chen, Wenjun Xie, Yuetong Luo (2020). "PointNGCNN: Deep Convolutional Networks on 3D Point Clouds with Neighborhood Graph Filters. *Computers and Graphics*, 86, 42-51. doi.org/10.1016/j.cag.2019.11.005.
- Ronggang Huang, Bisheng Yang, Fuxun Liang, Wenxia Dai, Jianping Li, Mao Tian, Wenxue Xu (2018). A Top-down Strategy for Buildings Extraction from Complex Urban Scenes Using Airborne LiDAR Point Clouds. *Infrared Physics and Technology*, 92, 203-18. doi.org/10.1016/j.infrared.2018.05.021.
- Wuzhao Li, Fu Dong Wang, Gui Song Xia (2020). A Geometry-Attentional Network for ALS Point Cloud Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 26-40. doi.org/10.1016/j.isprsjprs.2020.03.016.
- Xudong Lai, Yifei Yuan, Yongxu Li, Mingwei Wang (2019). Full-Waveform LiDAR Point Clouds Classification Based on Wavelet Support Vector Machine and Ensemble Learning. *Sensors*, 19(14). doi.org/10.3390/s19143191.
- Yangbin Lin, Cheng Wang, Dawei Zhai, Wei Li, Jonathan Li (2018). Toward Better Boundary Preserved Supervoxel Segmentation for 3D Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 143, 39-47. doi.org/10.1016/j.isprsjprs.2018.05.004
- Yi Yang, Hairong Fang, Yuefa Fang, Shijian Shi (2020). Three-Dimensional Point Cloud Data Subtle Feature Extraction Algorithm for Laser Scanning Measurement of Large-Scale Irregular Surface in Reverse Engineering. *Journal of Measurement*, 151, 107-220. doi.org/10.1016/j.measurement.2019.107220
- Yongguang Yang, Feng Chen, Fei Wu, Deliang Zeng, Yi-mu Ji, Xiao-Yuan Jing (2020). Multi-View Semantic Learning Network for Point Cloud-Based 3D Object Detection. *Neurocomputing*, doi.org/10.1016/j.neucom.2019.10.116.