TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

──────── \* ───────

**Báo Cáo Môn Học**

MÔN: PROJECT II

**Xây dựng 3D từ nhiều ảnh**

A close up of a sign

Description automatically generated

Sinh viên thực hiện : Trần Đức Anh Trường

Mã sinh viên : 20183648

Lớp : Khoa học máy tính 04

Giáo viên hướng dẫn : PGS.TS Đỗ Phan Thuận

***Hà Nội, tháng 07 năm 2021***

MỤC LỤC

[Lời Nói Đầu 3](#_Toc77425723)

[Phương Pháp Thực Hiện 4](#_Toc77425724)

[1. Xây dựng đồ thị tổng quát 4](#_Toc77425725)

[1.1. Ý tưởng. 4](#_Toc77425726)

[1.2. Các bước thực hiện. 4](#_Toc77425727)

[2. Phương pháp theo meshroom. 6](#_Toc77425728)

[2.1. Các khái niệm cơ bản. 6](#_Toc77425729)

[2.2. Các bước thực hiện: 9](#_Toc77425730)

[Kết Quả Đạt Được 11](#_Toc77425731)

[1. Xây dựng đồ thị tổng quát. 12](#_Toc77425732)

[2. Phương pháp theo meshroom. 12](#_Toc77425733)

[Kết Luận Và Định Hướng Phát Triển 15](#_Toc77425734)

[Tài Liệu Tham Khảo 16](#_Toc77425735)

# Lời Nói Đầu

Xây dựng 3D từ nhiều ảnh (stereophotogrammetry) là quá trình xây dựng mô hình 3 chiều từ nhiều ảnh 2 chiều.  Công nghệ này đang được ứng dụng rất rộng rãi và ngày càng được ưa thích, thay thế cho các phương pháp mô hình truyền thống do có nhiều ưu điểm vượt trội như tiết kiệm chi phí và thời gian, độ chi tiết, chính xác cao, có thể thực hiện được ở những nơi địa hình khó khăn, phức tạp…

Ảnh hoặc video có thể được quay chụp dễ dàng từ rất nhiều thiết bị khác nhau như điện thoại, máy ảnh, thiết bị bay (Drone, UAV)…

Phương pháp này được ứng dụng trong rất nhiều các lĩnh vực và mục đích khác nhau như: Khảo sát địa hình, địa vật trong xây dựng; Số hóa hiện vật cho bảo tàng ảo; Lưu trữ mô hình kiến trúc, lịch sử; Tạo mô hình phục vụ mô phỏng, làm phim, in 3D…

# Phương Pháp Thực Hiện

## Xây dựng đồ thị tổng quát

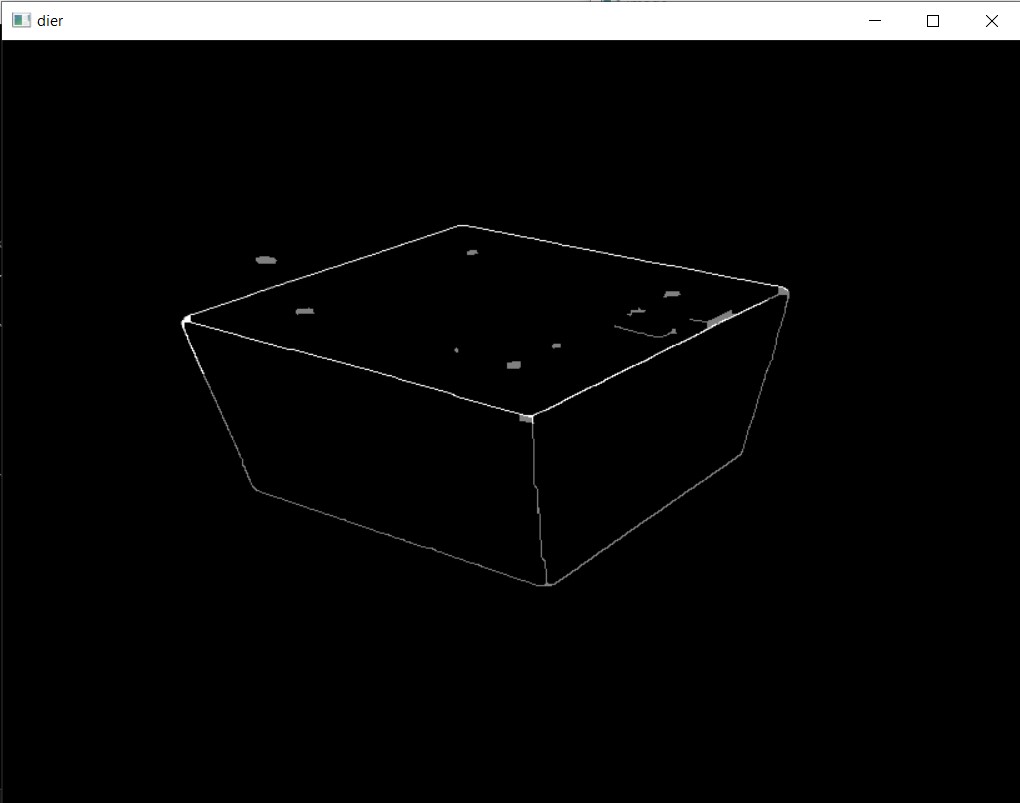
### Ý tưởng.

Mỗi 1 vật thể 3D thì ta có thể chuyển nó thành 1 đồ thị tổng quát dựa trên các tính chất và đặc trưng của nó. Ở mỗi góc nhìn thì ta có được 1 phần của vật thể cũng như 1 phần của đồ thị tổng quát. Nên từ mỗi ảnh thì ta trích xuất ra các đặc trưng và xây dựng ra 1 đồ thị để biểu diễn ảnh đó. Từ nhiều ảnh ở các góc khác nhau thì ta có các phần của đồ thị tổng quát. Sau đó kết hợp các phần lại thì ta có được đồ thị tổng quát. Từ đó xây dựng lên vật thể 3D.

### Các bước thực hiện.

1. Phát hiện vật thể:

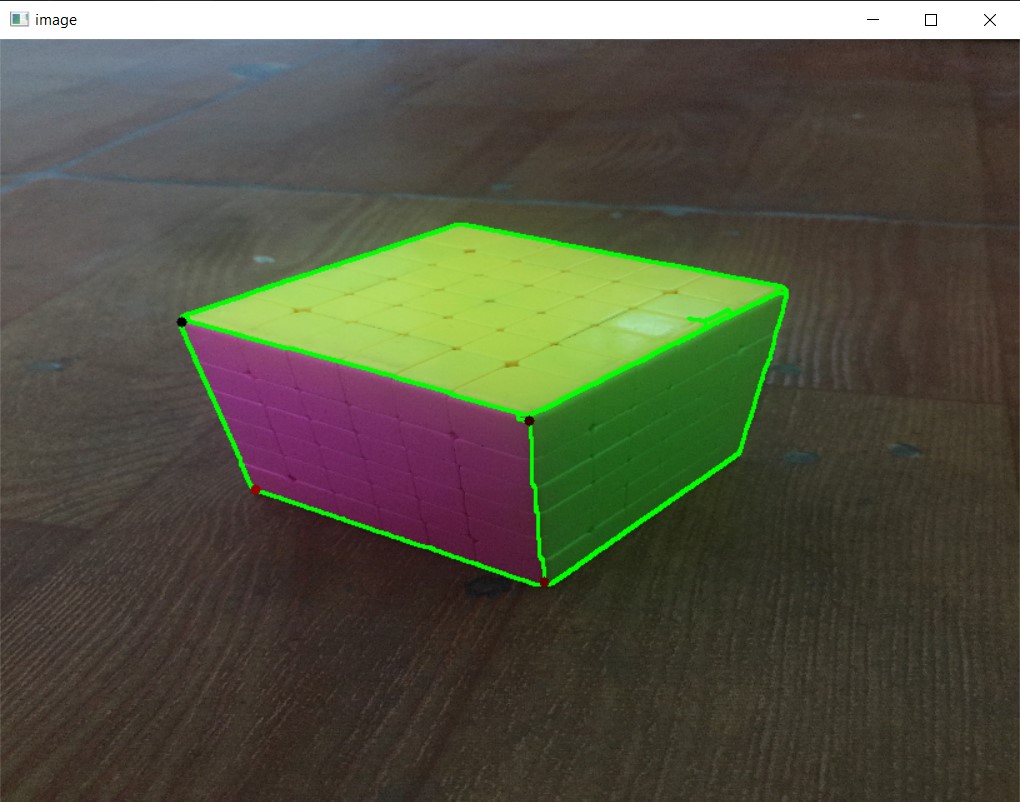
- Sử dụng 1 số cách làm mờ hình ảnh để lọc kết hợp với phát hiện cạnh (sử dụng Canny edge detect). Thu được ảnh chỉ chứa cạnh của vật thể.



Hình ảnh thu được sau khi phát hiện cạnh

1. Phát hiện các đặc điểm:

- Sử dụng Viền để có thông tin về các miền kết hợp phát hiện màu.



Viền đã phát hiện được

1. Mô hình hóa đỉnh:

- Từ các thông tin về miền (đỉnh, cạnh, màu,hình dạng, mối quan hệ giữa các cạnh, đỉnh) tạo thành 1 đỉnh.

1. Mô hình hóa cạnh:

- Từ mối quan hệ giữa các miền (có đỉnh, cạnh chung, vị trí gần nhau, tương quan màu sắc, ...) để tạo thành cạnh giữa các đỉnh.

1. Xây dựng đồ thị cho ảnh

- Tạo đồ thị riêng cho vật thể về các phần mặt xuất hiện trong ảnh.

1. Xây dựng đồ thị tổng quát

- Tổng hợp các đồ thị riêng thành 1 đồ thị tổng quát cho vật thể.

1. Xây dựng 3D

- Tái xây dựng các mặt (miền) từ các thông tin ở đỉnh

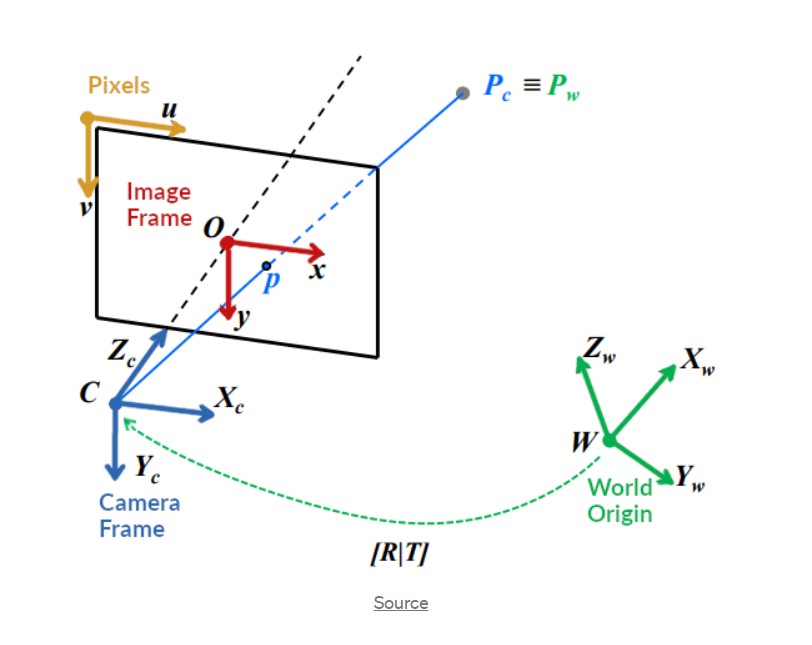
- Nối các mặt theo các cạnh đã có.

- Do phải thỏa mãn các rằng buộc (bất đẳng thức tam giác, khoảng cách cố định, …) về cạnh và đồ thị nên để không thể biểu diễn được trong 2D nên tạo cấu trúc 3D một cách tự nhiên.

## 2. Phương pháp theo meshroom.

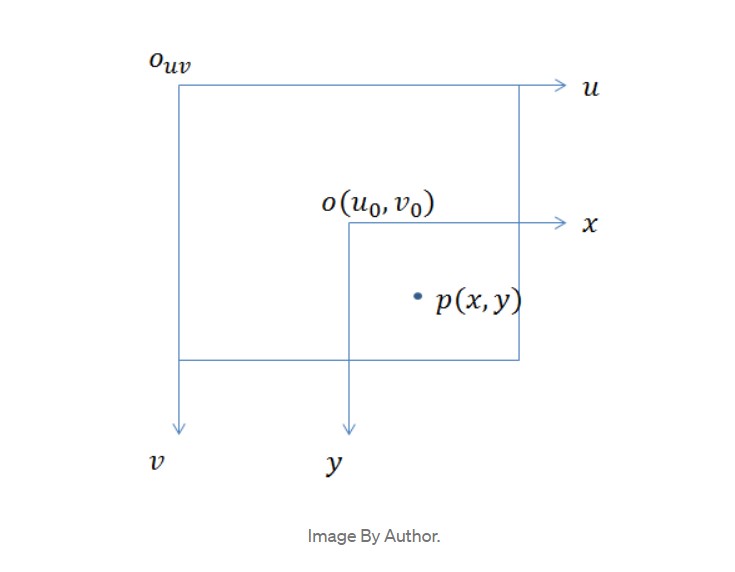
### 2.1. Các khái niệm cơ bản.

* Hệ tọa độ



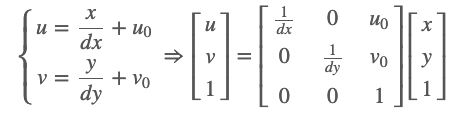
Structure from Motion

* Hệ tọa độ pixel ảnh (u, v)

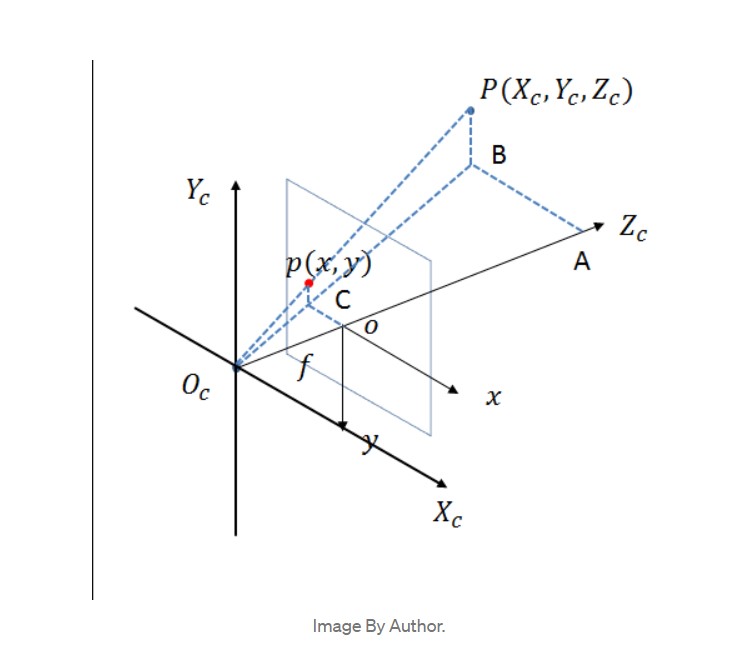


Structure from Motion

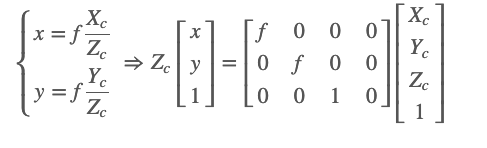
* Hệ tọa độ mặt phẳng ảnh (x, y)



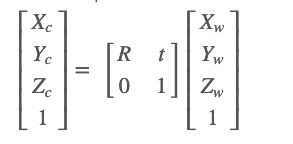
* Hệ tọa độ camera (Xc, Yc, Zc)



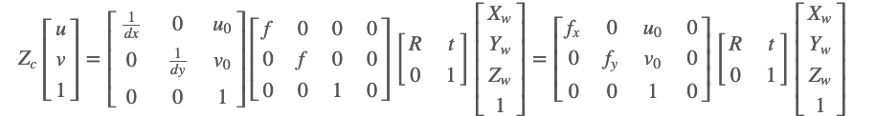
Structure from Motion



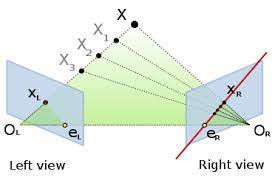
* Hệ tọa độ thực (Xw, Yw, Zw)



* Chuyển đổi hệ tọa độ



* Epipolar Geometry.



* + - Epipolar point: eL, eR
    - Epipolar line: eL xL, eR xR
    - Epipolar plane: OL X OR
* Ma trận
* E (ma trận cơ bản): ma trân mà x1c E x2c = 0 với x1c, x2c là điểm trong hệ tọa độ camera.
* K (ma trận nội tại): đặc trưng cho camera.
* [R|T] (ma trận bên ngoài): ma trân biến đổi tọa độ camera sang tọa độ thực.

Với R là ma trận quay, T là ma trận tịnh tiến

* Rằng buộc cơ bản

Trong hệ tọa độ homogeneous:

* + x: tọa độ của điểm trong ảnh (u, v, 1)T.
  + X: tọa độ của vật thể trong không gian (Xw, Yw, Zw, 1)T

Từ quan hệ giữa các hệ tọa độ ta được:

s.x = K(R.X + T) (với s là 1 biến số)

=> s.K-1x = R.X + T

Đặt K-1x = x’ là tọa độ của điểm trên ảnh trong không gian

* s.x’ = R.X + T

Trong hệ tọa độ thực tùy ý nên với hệ có 2 điểm ta chọn điểm điểm gốc tại camera bất kỳ.

s1.x’1 = X

s2.x’2 = R.X + T

=> s2.x’2 = R. s1.x’1 + T

Vì x2’ vuông góc T

* s1.(T\*.x2’)T.R. x’1 = 0
* (x2’)T.T\*.R. x’1 = 0

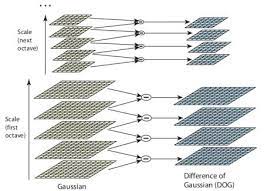
Đặt E = T\*.R

* (x2’)T.E.x’1 = 0 (Rằng buộc cơ bản)

### 2.2. Các bước thực hiện:

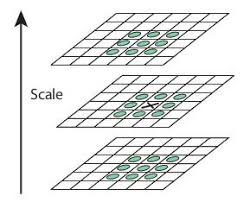
a) Khai thác đặc trưng tự nhiên

* Phát hiện các đặc trưng của ảnh mà không thay đổi khi bị méo, xoay.
* Phát hiện các đặc trưng theo SIFT [6].
* Phát hiện các cực trị không gian theo tỉ lệ:
  + Làm mờ Gauss theo biến và k
  + Tính sự khác nhau giữa các lớp Gauss



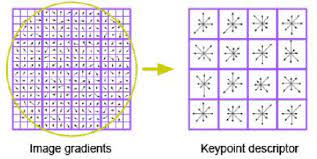
Distinctive Image Features from Scale-Invariant Keypoints

* + So sánh 1 điểm trên ảnh với 26 điểm xung quanh. Nếu nó nhỏ hơn hoặc lớn hơn tất cả thì đó là 1 điểm chính tiềm năng



Distinctive Image Features from Scale-Invariant Keypoints

* Cục bộ hóa điểm chính
  + - Loại bỏ các điểm chính không tốt.
* Gán hướng
  + - Gán hướng và độ lớn cho điểm chính.
* Bộ mô tả điểm chính



* Gán hướng cho các điểm xung quanh điểm chính và tạo nên 1 bộ mô tả điểm chính thành 128 (16 x 8) ngăn.

b) Ghép ảnh

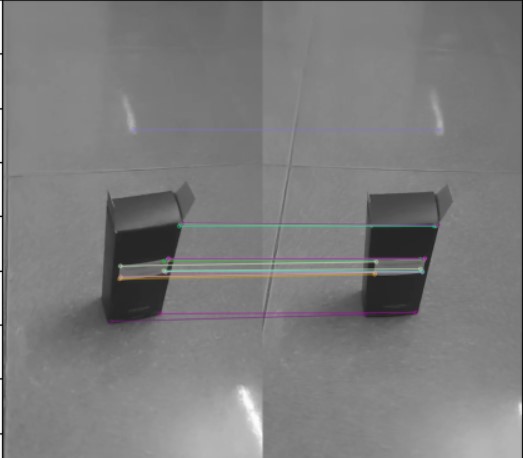
- Tìm các ảnh có cùng hình ảnh về 1 phần của vật thể.

* Cắt ảnh từ video theo cặp thành một chỗi các ảnh theo thứ tự.
* Có thể tham khảo phương pháp the vocabulary tree [7].

c) Ghép đặc trưng

- Tìm các đặc trưng giống nhau cùng trỏ về 1 vị trí trong không gian

* Ghép cực hạn để chính xác nhất có thể
  + Tìm kiếm tất cả các bộ mô tả điểm chính và chọn ra bộ giống nhất



Ghép đặc trưng

* Có thể tham khảo Tìm kiếm hàng xóm gần nhất, KDTree.

1. Xây dựng cấu trúc từ chuyển động [5]

* Xây dựng đám mây điểm về vật thể từ chuyển động.
* Hiệu chỉnh camera: Tính toán ma trận nội tại (K) của camera thông qua tập ảnh bàn cờ vua. Đã có 1 phương pháp có sẵn trong OpenCV để tính ra ma trận nội tại của camera cho ảnh có bàn cờ vua.
* Tính toán ma trận bên ngoài (E) dùng 8-point algorithm:
  + Hai ảnh được tính toán cần có 8 bộ mô tả điểm chính giống nhau.
* Tính toán ma trận quay (R), ma trận tịnh tiến (T) bằng cách phân tách E theo SVD phân tách.
* Tính toán tọa độ 3D của vật thể từ 2 điểm 2D bằng trigulation (thị sai). [9]

1. Ước lượng độ sâu [8]

* Đánh giá độ sâu của vật thể.
* Dựa vào độ sâu của vật thể ta có thể tính được khoảng cách tương đối với các vật xung quang và đo được khoảng cách chính xác hơn.
* Tính toán theo cặp ảnh áp dụng disparity theo OpenCV.

f) Nối lưới

* Ghép các điểm của đám mây điểm thành 1 lưới điểm tạo thành hình dạng của vật thể.

g) Tạo kết cấu

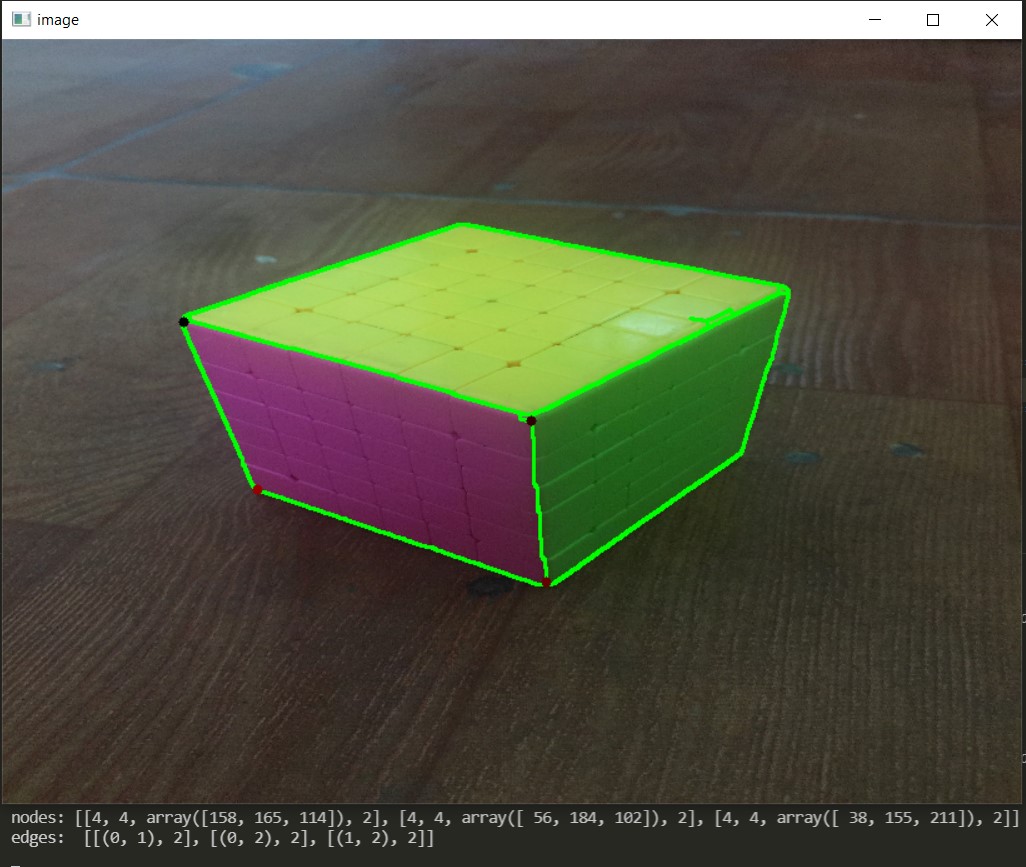
* Tạo kết cấu vật thể cho lưới đã tạo thành 1 vật 3D hoàn chỉnh.

h) Cục bộ hóa

* Hiệu chỉnh camera.

# Kết Quả Đạt Được

## Xây dựng đồ thị tổng quát.



Kết quả

Xây dựng đồ thị của ảnh theo thời gian thực.

Kết quả cho thấy:

* Phân tích được các đỉnh và cạnh của đồ thị
* Các đỉnh của đồ thị chứa thông tin chính xác về số đỉnh, số cạnh, số liên kết tới đỉnh khác
* Các cạnh chứa thông tin chính xác về các đỉnh đã liên kết

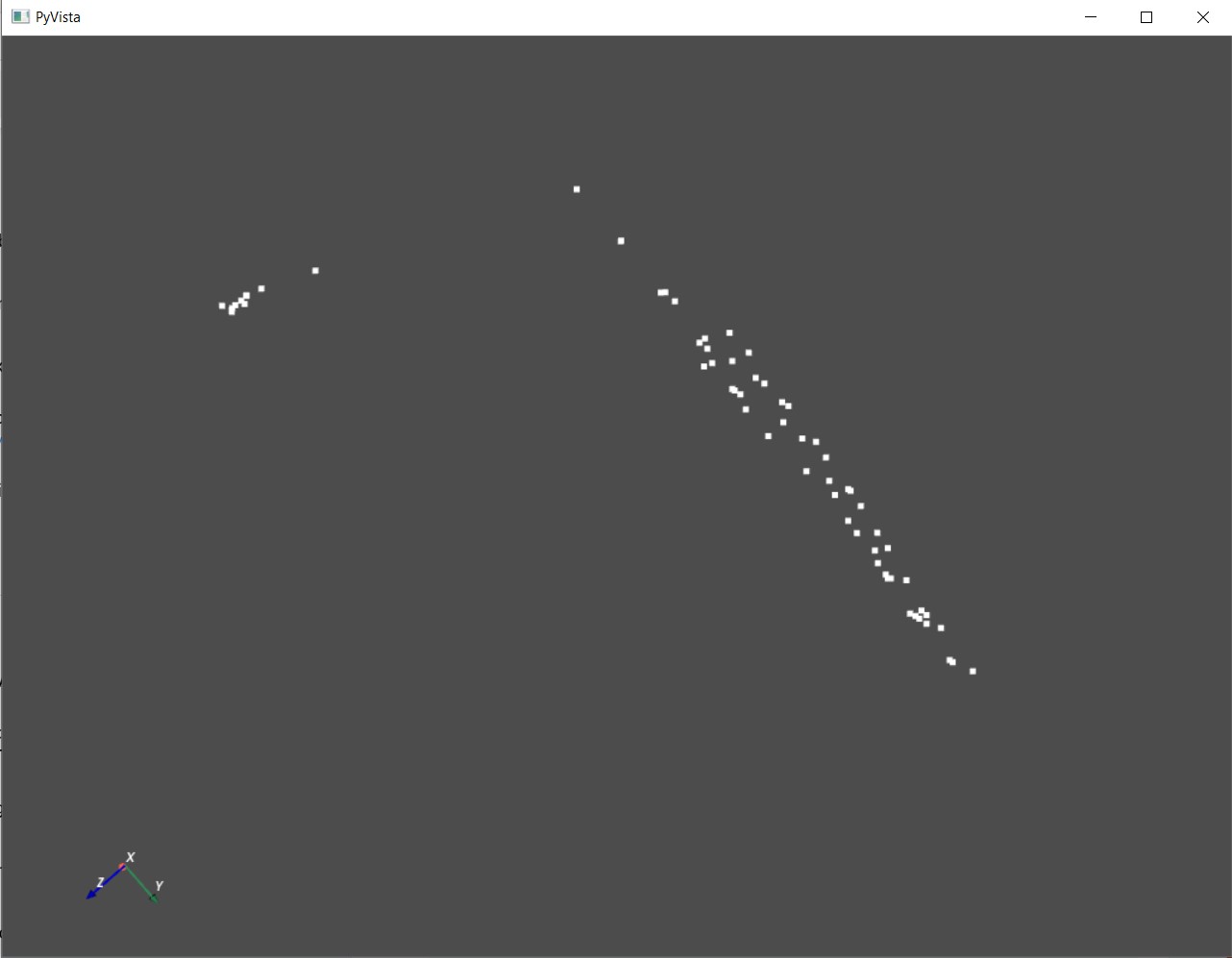
## 2. Phương pháp theo meshroom.

Đạt được đám mây điểm của vật thể trong 1 số trường hợp nhưng chưa tốt.

Kết quả đám mây điểm của 1 vật thể



Ảnh của vật thể







Đám mây điểm của vật thể

Kết quả cho thấy đã tạo được 1 đám mây điểm của 1 mặt và phần mặt trên của vật thể.

# Kết Luận Và Định Hướng Phát Triển

Sau khi tìm hiểu và áp dụng các phương pháp của mình và meshroom thì đã có một vài kết quả nhưng chưa được tốt và chính xác.

Tóm lại các mục đã trình bày, em đã học được trong quá trình làm Project 2:

* Tìm hiểu thêm nhiều về xử lý ảnh.
* Tìm hiểu các định nghĩa và thuật toán cơ bản trong phép đo lập thể (stereophotogrammetry).
* Áp dụng các thuật toán và sử dụng với OpenCV.
* Có được kết quả nhưng chưa được tốt.

Có 1 số phần mềm mã nguồn mở và được áp dụng trong thực tế

* ODM (Open Drone Map) dùng cho ảnh chiếu xuống từ trên cao.
* Meshroom dùng cho các vật thể đơn lẻ.

Định hướng phát triển theo cách xây dựng đồ thì tổng quát.

* + Tích hợp được các thông tin về bề mặt vào các đỉnh của đồ thị dễ dàng.
  + Xây dựng một cách có cấu trúc.
  + Tùy biến một cách dễ dàng và thực tế hơn.
  + Kết quả đầu ra không cần quá chính xác mà cần tập chung vào cấu trúc của vật thể.

# Tài Liệu Tham Khảo

[1] https://alicevision.org/#photogrammetry

[2] https://docs.opencv.org/master/d6/d00/tutorial\_py\_root.html

[3] <https://github.com/alicevision/meshroom>

[4] Wikipedia.org

[5] https://towardsdatascience.com/structure-from-motion-311c0cb50e8d

[6] Distinctive Image Features from Scale-Invariant Keypoints

[7] Scalable Recognition with a Vocabulary Tree, David Nister and Henrik Stewenius, CVPR 2006

[8] Accurate and efficient stereo processing by semi-global matching and mutual information, H. Hirschmüller. CVPR 2005

[9] https://en.wikipedia.org/wiki/Triangulation\_(computer\_vision)

## 