ĐẠI HỌC QUỐC GIA HÀ NỘI

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Khoa Điện tử – Viễn thông

BÁO CÁO CUỐI KỲ

Đề tài:

3D OBJECT LOCALIZATION

GVHD:	**Hoàng Văn Xiêm, Đỗ Ngọc Minh**
Bộ môn:	**Xử lý ảnh và thị giác Robot**
Ngành:	**Kỹ thuật Robot**
Sinh viên:	**Phạm Thị Mỹ Lệ**
MSV:	**21020445**

Hà Nội, ngày 24 tháng 12 năm 2023

Mục lục

Danh mục hình ảnh

Chương 1: Tổng quan

1. Giới thiệu bài toán

Trong bối cảnh ngày càng tăng cường của tự động hóa và robot hóa trong các ứng dụng công nghiệp và dịch vụ, khả năng của robot trong việc nhận diện và tương tác với môi trường ngày càng trở nên quan trọng. Một trong những thách thức lớn là khả năng của robot trong việc nhìn và hiểu rõ đối tượng một cách chi tiết từ hình ảnh 2D. Bài toán chuyển đổi tọa độ ảnh 2D sang tọa độ ở trạng thái 3D trở thành một khía cạnh quan trọng để nâng cao khả năng tương tác của robot, đặc biệt là trong quá trình gắp vật ở môi trường sản xuất công nghiệp. Thay vì dựa vào thông tin hình ảnh 2D truyền thống, chúng ta cần một quy trình chuyển đổi thông tin đó thành không gian 3D, giúp robot có khả năng hiểu biết về hình dạng, kích thước và vị trí của vật thể trong không gian. Ứng dụng này không chỉ giúp robot chọn lựa vật thể cần gắp một cách chính xác mà còn cải thiện khả năng xử lý và tương tác trong môi trường đa dạng. Sự kết hợp giữa hình ảnh 2D và dữ liệu 3D cung cấp cho robot cái nhìn đa chiều, mở ra khả năng tối ưu hóa quá trình gắp vật, tránh va chạm không mong muốn và tăng cường độ an toàn trong các ứng dụng công nghiệp, logictics, và dịch vụ. Đồng thời, việc áp dụng công nghệ này không chỉ làm tăng cường khả năng của robot mà còn mở ra những tiềm năng mới trong lĩnh vực tự động hóa và dịch vụ thông minh.

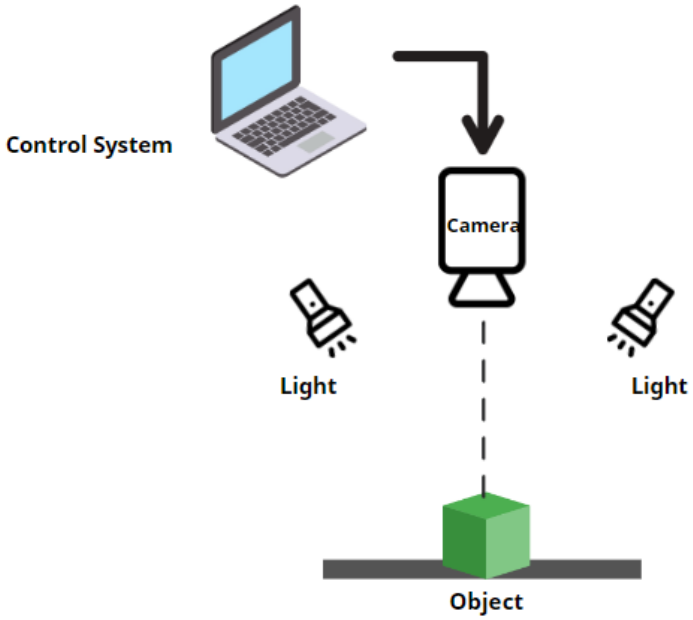
2. Các nghiên cứu liên quan

Với chủ đề chuyển đổi tọa độ từ 2D sang 3D ở thế giới thực, có rất nhiều các nghiên cứu liên quan. Một trong những phương pháp phổ biến để thực hiện việc chuyển đổi này là sử dụng mạng Nơ-ron, đặc biệt là CNN, do nó có thể xử lý dữ liệu đa chiều và tự động hóa quá trình học. Nhưng bên cạnh đó đòi hỏi phải có một lượng lớn dữ liệu để có thể đạt được hiệu suất tốt và tài nguyên tính toán phức tạp.

Do đó, với yêu cầu đề bài đặt ra với môn học và các khả năng cho phép, bài báo cáo này sẽ được đi theo hướng thứ 2, sử dụng lý thuyết tính toán của giải tích và đại số tuyến tính và hình học để tìm ra được các quy luật giữa các đối tượng cần xét tới để có thể giải quyết được bài toán. Mặc dù, với phương pháp này có vẻ truyền thống hơn nhưng nó lại có các ưu điểm về tính toán và không yêu cầu đòi hỏi lượng dữ liệu lớn mà vẫn đạt được hiệu suất tốt, linh hoạt được trong các bài toán cụ thể.

Chương 2: Mô hình, Lý thuyết và Thuật toán

1. Mô phỏng hệ thống



Hình 1: Mô phỏng hệ thống

Bài toán đưa ra với mục đích có thể ứng dụng trong công nghiệp, nên hệ thống sẽ bao gồm 1 camera được chiếu thẳng đứng từ trên xuống với khoảng cách là không đổi, một vật 3D vuông góc với camera và có hệ thống chiếu sáng đầy đủ để đạt được hiệu suất cao nhất.

2. Lý thuyết

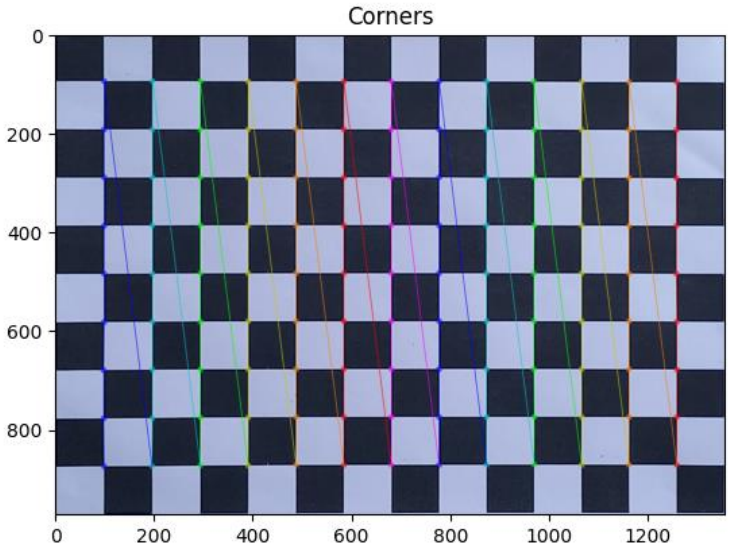
Hiệu chỉnh máy ảnh là không thể thiếu trong các ứng dụng thị giác máy tính vì hầu hết các hệ thống thị giác máy tính đều bị ảnh hưởng nhiều bởi độ chính xác của việc hiệu chuẩn. Hiệu chuẩn liên quan đến những phát triển chính trong quá trình tái tạo 3D bao gồm khôi phục hình học của máy ảnh, trích xuất và phân tích thông tin 3D. Kỹ thuật này cũng được sử dụng để ước tính vị trí 3D và góc quay của camera so với các thông số bên ngoài và bên trong, cung cấp thông tin về tọa độ thế giới 3D và thể hiện các đặc tính quang học của camera tương ứng. Việc xác định các tham số của hàm để giải thích ánh xạ từ vị trí của một điểm trong tọa độ 3D đến vị trí của một điểm trên mặt phẳng ảnh là một trong những mục tiêu chính của việc hiệu chỉnh máy ảnh. Các tham số hình học, còn được gọi là tham số camera, xác định theo kinh nghiệm mối quan hệ giữa vị trí camera và tọa độ thế giới 3D.

Bên cạnh đó việc sử dụng bàn cờ trong hiệu chỉnh máy ảnh, cho phép các thông số máy ảnh trích xuất thông tin chính xác hơn từ hình ảnh.

Có 2 thông số chính trong máy ảnh:

- Thông số nội tại (thông số bên trong máy ảnh) cung cấp các đặc điểm hình học và quang học của máy ảnh bao gồm độ dài tiêu cự, tâm hình ảnh và độ biến dạng của ống kính
- Thông số bên ngoài máy ảnh: cung cấp hướng và vị trí 3D của máy ảnh liên quan đến tọa độ thế giới

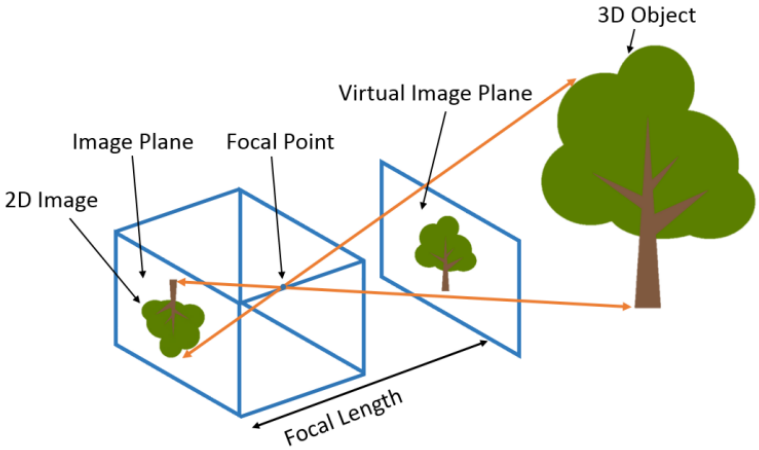
Trong bài báo cáo này, máy ảnh sẽ được hiệu chỉnh với với bàn cờ có kích thước là 13x9, với mỗi ô khoảng 17.5mm.



Hình 2: Calibration chessboard

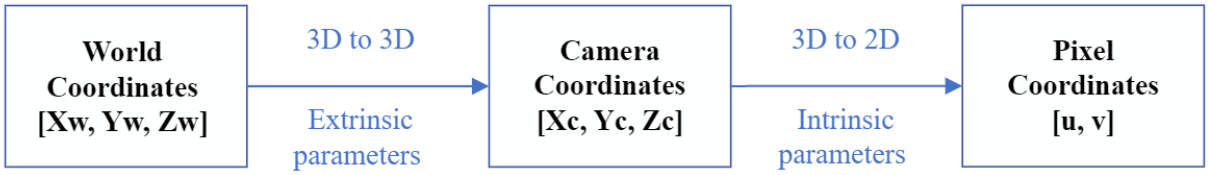
3. Toán học về ánh xạ tọa độ điểm ảnh từ không gian 2D sang không gian 3D

Khi một vật thể 3D đi qua tia sáng, mối quan hệ của tọa độ vật và tọa độ ảnh 2D của phép chiếu sẽ được mô tả thông qua hình ảnh về pinhole [1] camera phía dưới:



Hình 3: Pinhole Camera

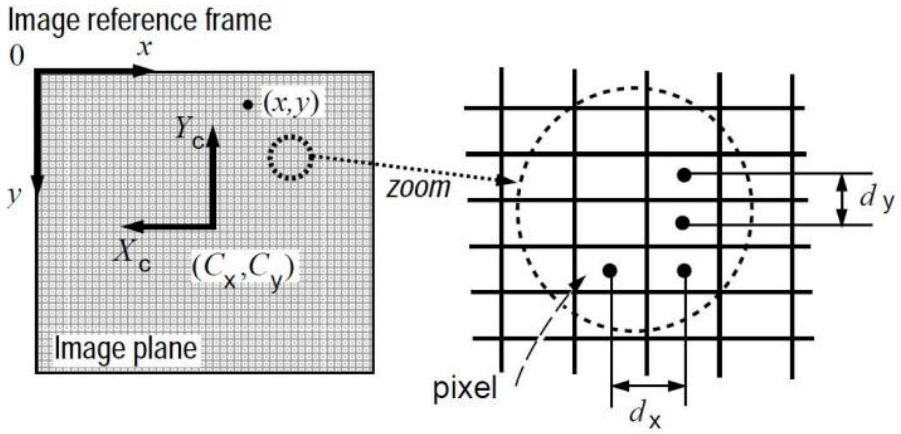
Giả sử đã có tọa độ 2D điểm ảnh, mục đích của ta là cần tìm tọa độ 3D của vật ở ngoài thực tế. Để có thể làm được điều đó, chúng ta cần phải ánh xạ tọa độ điểm ảnh sang tọa độ camera, sau đó ánh xạ từ hệ camera sang hệ tọa độ thế giới.



Hình 4: Sơ đồ ánh xạ

- **Chuyển đổi từ hệ 2D pixel sang 2D mm**

Đầu tiên, ta xét tọa độ điểm ảnh (2D) có tọa độ là (x,y) với đơn vị là pixel. Vì các tọa độ trên hệ camera và hệ thế giới được tính theo đơn vị mm, ta cần phải chuyển đổi từ hệ pixel sang hệ mm. Ta chọn gốc của khung tham như hình vẽ.



Hình 5: Sơ đồ chuyển đổi giữa hệ pixel và minimet

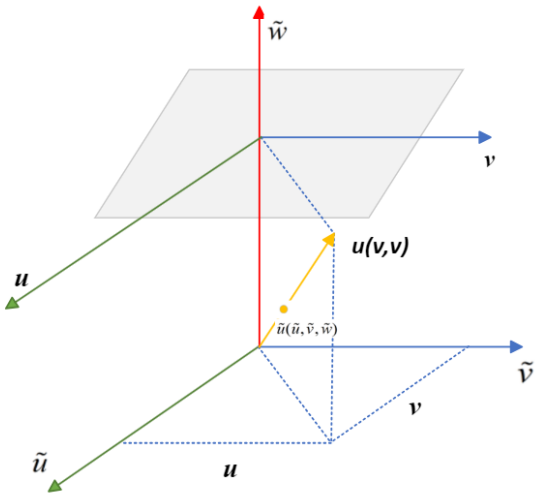
Từ đó, ta quy đổi được tọa độ của một điểm trong hệ quy chiếu dựa trên pixel:

$$u = m_x . f . \frac{x_c}{z_c} = f_x . \frac{x_c}{z_c}$$

$$v = m_y . f . \frac{y_c}{z_c} = f_y . \frac{y_c}{z_c}$$; mx, my là mật độ pixel theo trục x, y; f là tiêu cự máy ảnh

- **Chuyển đổi từ hệ pixel sang camera**

Một điểm 2D được ký hiệu là $m = [u, v]^T$. Và điểm 3D ánh xạ trên hệ camera là (Xc, Yc, Zc). Do 2 vecto khác nhau về chiều, nên chúng ta sẽ đồng nhất hệ tọa độ mặt phẳng bằng cách tăng cường bằng cách thêm 1 làm phần tử cuối cùng của vecto 2D: $m_e = [u, v, 1]^T$



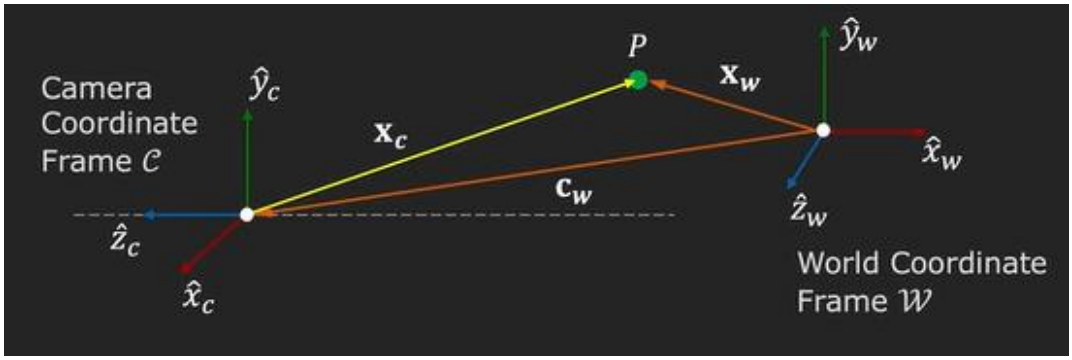
Hình 6: Chuyển đổi từ hệ 2D sang hệ camera

Đồng nhất tọa độ ta thu được công thức sau:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} z_c.u \\ z_c.v \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & O_x & 0 \\ 0 & f_y & O_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

Đặt ma trận thông số nội tại của máy ảnh: $K = \begin{bmatrix} f_x & 0 & O_x \\ 0 & f_y & O_y \\ 0 & 0 & 1 \end{bmatrix}$ → **Intrinsic Matrix**

Để chuyển đổi từ tọa độ camera sang tọa độ thực, vị trí mà chúng ta bàn tới sẽ bao gồm vector quay và Ma trận xoay



Hình 7: Chuyển đổi từ hệ Camera sang hệ thế giới thực

Trong đó ma trận xoay có dạng: $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$

Với các hàng đại diện với hướng theo tứ tự xc yc zc trong tọa độ thế giới

Từ hệ tọa độ trên, ta có thể suy ra $X_c = R.(X_w - C_w) = R.X_w + t$

$$\text{Hay: } X_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R. \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Đồng nhất hệ tọa độ ta thu được công thức như sau:

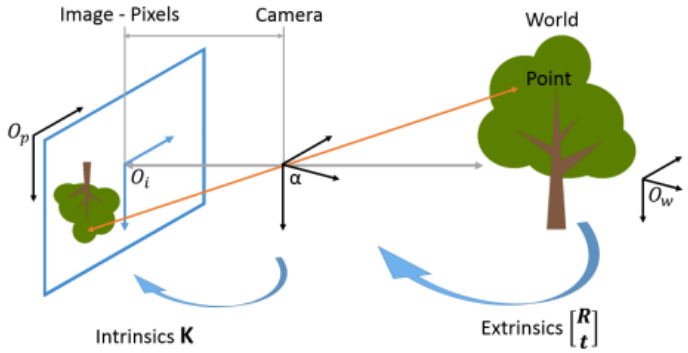
$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \rightarrow \textbf{Extrinsic Matrix}$$

Ma trận 4x4 trên còn được gọi là ma trận ngoài camera. Để chuyển từ ma trận thế giới sang trận camera, ta sẽ thông qua ma trận Extrinsic

Do đó, tổng quát để chuyển từ hệ tọa độ thực (3D) sang hệ tọa độ ảnh (2D):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P. \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \rightarrow P = K.[R \mid t]$$

Trong đó R và t lần lượt là các phần tử quay và tịnh tiến của ma trận bên ngoài, trong khi K đại diện cho ma trận bên trong. Các tham số bên ngoài chuyển đổi tọa độ 3D thành tọa độ camera và sau đó các tham số bên trong chuyển đổi tọa độ camera thành mặt phẳng hình ảnh như hình dưới đây:

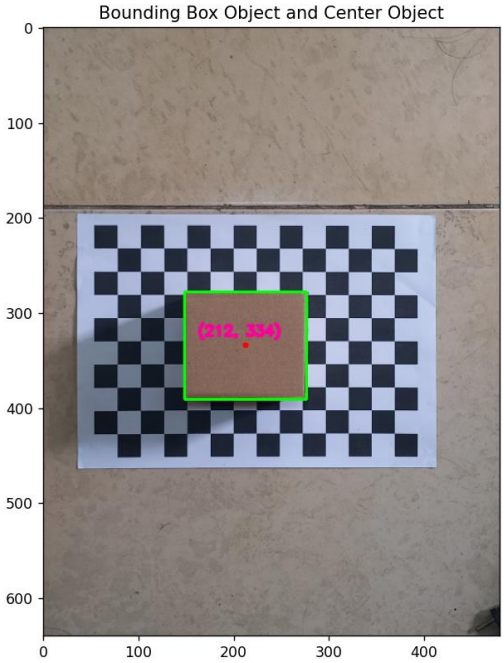


Hình 8: Tổng quan ma trận Camera

Chương 3: Kết quả và đánh giá

1. **Quá trình tìm tọa độ tâm của vật trên ảnh 2D:**

   Được thực hiện bằng cách xét ngưỡng màu của vật, và tìm các cạnh biên của vật tương ứng. Sau đó ta bounding box vật và tìm tọa độ trọng tâm dựa vào hình bao.



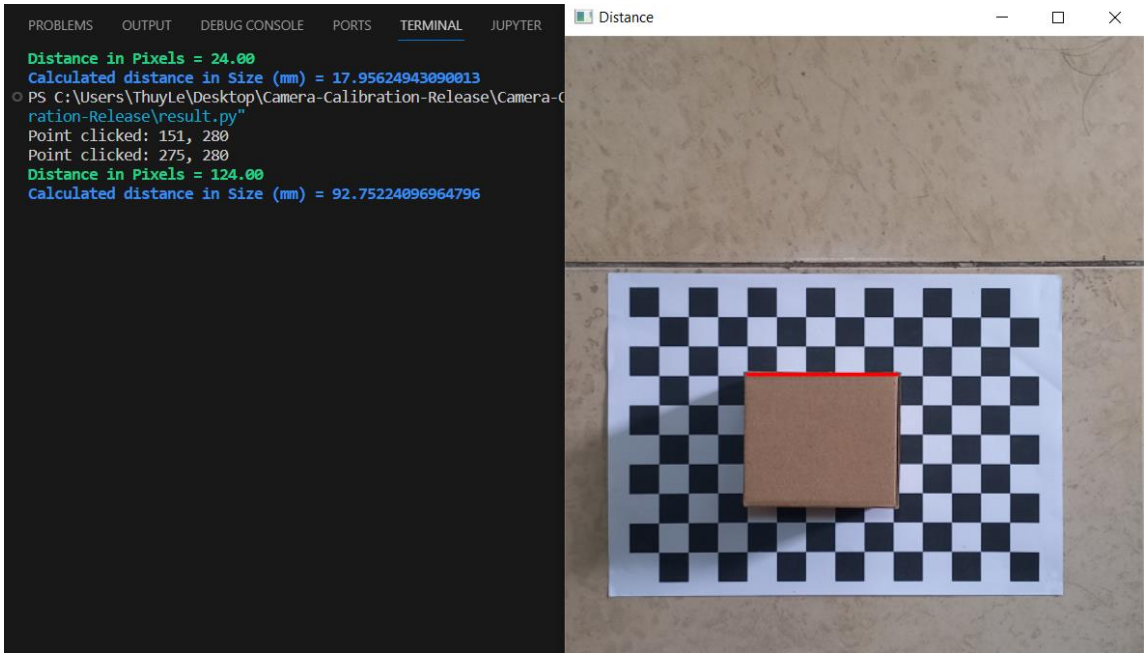
Hình 9: Kết quả Bounding box & center object

➔ **Nhận xét:** Bởi vì có dùng tới ngưỡng màu của vật, nên bài toán chỉ dùng trong một số trường hợp cụ thể, và ánh sáng cố định. Nếu ánh sáng bị thay đổi, sẽ ảnh hưởng tới kết quả bài ra và tọa độ tâm tìm được

2. **Ánh xạ tọa độ 2D sang tọa độ 3D**

   Sử dụng các kiến thức toán thuần, ta sẽ tính được các thông số trong và ngoài của camera. Từ đó ta sử dụng để có thể đưa vào trong quá trình tính toán kết quả đầu ra.

   Vì trong quá trình đo đạc trọng tâm ở thế giới thực có thể xảy ra sai số, trước tiên ta tiến hành đo độ dài cạnh của vật để kiểm gia sai số

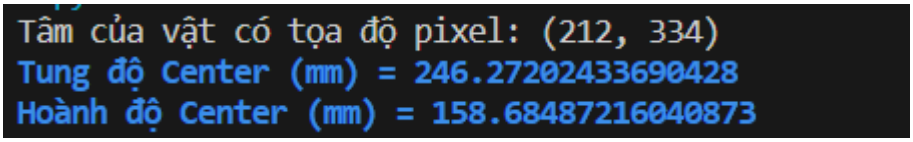


Hình 10: Kết quả kiểm thử

➔ **Nhận xét:** Kết quả nhận được là 92.75mm, trong khi số đo thực tế là 91mm

➔ sai số chưa tới 2%

Khi đã có tọa độ tâm trong ảnh 2D (pixel), ta tiến hành nhân nghịch đảo với các ma trận K, R để tìm ra tọa độ trong thế giới thực 3D



Hình 11: Thực nghiệm kết quả tọa độ vật 3D

Sau khi tiến hành một loạt các data cho trước, ta nhận thấy rằng, sai số không quá 5%

Chương 4: Kết luận

Như vậy, trong bài báo cáo này đã tập trung vào việc trình bày các kỹ thuật định vị đối tượng dựa trên ước tính tư thế và hiệu chỉnh máy ảnh. Chỉ thông qua một góc nhìn chính diện từ trên xuống thông qua máy ảnh, ta có thể ước tính được tọa độ thực của vật thể 3D theo 2 chiều x và y. Kết quả của các tham số camera và vị trí đối tượng trong không gian 3D được ước tính một cách khá chính xác, với sai số trong quá trình ở khoảng 5%.

Bên cạnh đó, ta có thể rút ra kết luận được rằng nếu chúng ta sử dụng nhiều hình ảnh 2D từ các góc độ khác nhau, kết hợp với các thuật toán ước tính tư thế, có thể dẫn đến việc tái tạo được tọa độ 3D của đối tượng một cách hiệu quả.

Tài liệu tham khảo

[1] T. H. M. S. a. M. Usman, "3D Object Localization Using 2D Estimates for," 2020.

[2] S. Peng∗, "Calibration Wizard: A Guidance System for Camera Calibration".

[3] G.-Q. W. a. S. Ma, "A complete two-plane camera calibration method and experimental comparisons," ICCV, 1993.

[4] H. C. a. F. Tsai, " Vanishing point extraction and refinement for Robus," 2018.

[5] C. 7. S. –. 2012, "Camera Calibration," 2012.

[6] B. Bhatt, "Analytics Vidhya," What is Camera Calibration in Computer Vision?, 13 6 2013. [Online].

[7] F. P. o. C. Vision, Composer, *Linear Camera Model | Camera Calibration.* [Sound Recording]. 2021.

[8] P. S. Songyou Peng, "Calibration Wizard: A Guidance System for Camera Calibration," arXiv:1811.03264v1, 2018.

[9] Z. Zhang, "A Flexible New Technique for Camera," 2008.

[10] R. Staszak and D. Belter, "3D Object Localization With 2D Object Detector and 2D Localization," 2022.

[11] S. Mohammed, M. Z. A. Razak and A. H. A. Rahman, "An Efficient Intersection Over Union Loss Function for 3D Object Detection".