Xác định tọa độ tâm của vật thông qua Camera điện thoại

Nguyễn Hải Thành – 20020717
Bộ môn Kỹ thuật Robot, Khoa Điện tử - Viễn thông
Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

I. GIỚI THIỆU

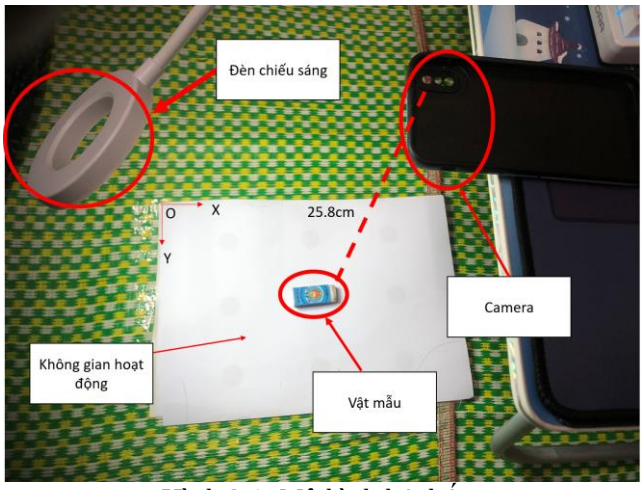
Trong thời đại ngày nay, xử lý ảnh đã trở thành một lĩnh vực quan trọng với nhiều ứng dụng đa dạng trong các lĩnh vực như công nghiệp, y tế, xe tự động và nhiều lĩnh vực khác. Một trong những thách thức quan trọng là xác định vị trí tâm của vật thể trên ảnh, có thể là ảnh 2D hoặc ảnh 3D từ các thiết bị như camera, cảm biến laser, hoặc thiết bị chụp hình 3D.

Trong việc xác định tâm 2D của vật thể trên ảnh, các thuật toán xử lý ảnh như Canny, contours detection, hoặc các phương pháp máy học như Học Sâu (Deep Learning) thường được sử dụng. Các phương pháp này giúp nhận diện đường viền và đối tượng trong ảnh, từ đó xác định được tâm của chúng dựa trên thông tin hình học.

Khi xử lý ảnh 3D, thông tin chiều sâu của vật thể trở thành một yếu tố quan trọng. Các thiết bị như camera 3D, lidar, hoặc cảm biến toàn cảnh 3D cung cấp thông tin về khoảng cách từ vật thể đến máy ảnh. Bằng cách sử dụng thông tin này, các thuật toán có thể xác định tâm 3D của vật thể trong không gian.

Xác định tâm của vật thể trong ảnh có ứng dụng rộng rãi trong nhiều lĩnh vực. Trong công nghiệp, nó có thể được sử dụng để theo dõi sản phẩm trên dây chuyền sản xuất. Trong y tế, nó có thể giúp xác định vị trí của các cơ quan trong cơ thể. Trong xe tự động, xác định tâm là quan trọng để theo dõi và tránh va chạm.

II. MÔ TẢ HỆ THỐNG



Hình 2.1. Mô hình hệ thống

Nghiên cứu này tập trung vào việc phát triển một hệ thống xác định tọa độ tâm của vật thể từ hình ảnh thu được từ camera điện thoại iPhone X, được đặt cố định ở khoảng cách 25.8cm từ mặt phẳng hoạt động đến vật thể quan sát. Môi trường làm việc được giới hạn trong kích thước của một tờ giấy A4 (297x210mm), và hệ tọa độ được thiết lập tại góc trái trên của tờ giấy.

Hệ thống được cải tiến bằng cách tích hợp đèn chiếu sáng có khả năng tăng giảm độ sáng. Đèn này không chỉ làm nổi bật chi tiết của vật thể mà còn tối ưu hóa chất lượng hình ảnh thu được trong các điều kiện ánh sáng biến động. Đặc biệt, đèn chiếu sáng có khả năng điều chỉnh giúp hệ thống linh hoạt hơn trong việc xử lý các điều kiện ánh sáng khác nhau.

Phần cứng của hệ thống bao gồm camera iPhone X với cảm biến chất lượng cao, và đèn chiếu sáng được tích hợp. Phần mềm của hệ thống sử dụng các thuật toán xử lý hình ảnh phức tạp để xác định tọa độ tâm của vật thể trong không gian 2D. Kết hợp giữa camera và đèn chiếu sáng, hệ thống không chỉ cung cấp khả năng xác định tọa độ mà còn đảm bảo độ chính xác và đáng tin cậy trong nhiều điều kiện ánh sáng, mở ra nhiều ứng dụng trong theo dõi và giám sát, thực tế ảo, và công nghệ thị giác máy tính.

III. PHƯƠNG PHÁP GIẢI QUYẾT

Phương pháp giải quyết vấn đề trong quá trình hiệu chỉnh máy ảnh và tái tạo 3D đòi hỏi một sự kết hợp cân đối giữa hiểu biết sâu sắc về lý thuyết và sự sáng tạo trong việc đối mặt với thách thức cụ thể. Đầu tiên và quan trọng nhất, tôi đã chú trọng đến việc hiểu rõ các nguyên lý cơ bản của hiệu chỉnh máy ảnh, bao gồm cả ma trận camera, hệ số biến dạng, và thông số ngoại tại.

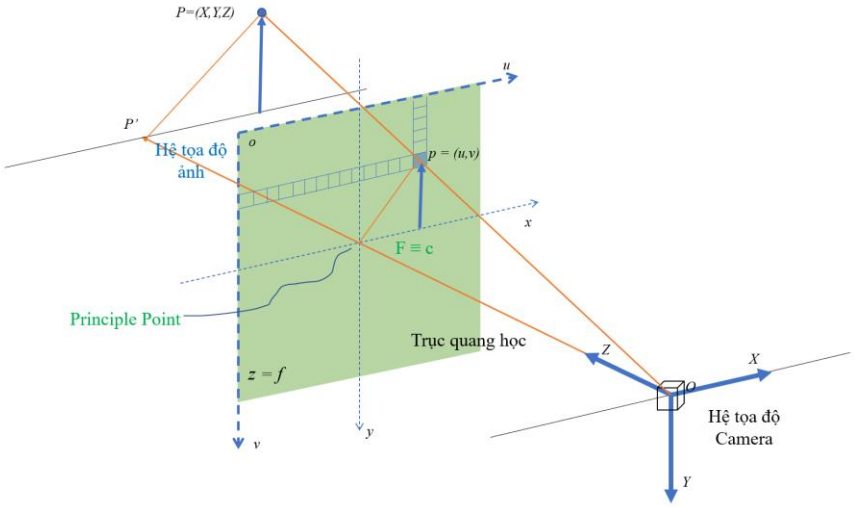
A. Hiệu chỉnh camera và tái tạo 3D

Một trong những vấn đề lớn xuất phát từ mô hình là tính toán tọa độ XYZ trong thế giới thực từ các điểm chiếu của Hình ảnh nhất định.

$$\underbrace{s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}}_{\text{Given this}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\text{Find this}}$$

- (X, Y, Z) là tọa độ của một điểm 3D trong không gian thế giới.
- (u, v) là tọa độ của điểm chiếu trong pixel.
- A là ma trận camera, hoặc ma trận các tham số nội tại.
- (c_x, c_y) là điểm chính là điểm thường nằm ở tâm ảnh.
- f_x, f_y là tiêu cự được biểu thị bằng đơn vị pixel.

Vì không thể tính nghịch đảo của ma trận R|t vì nó không phải là ma trận vuông. Để vượt qua vấn đề này, tôi đã áp dụng một phương pháp linh hoạt bằng cách thêm hệ số tỷ lệ và ma trận camera vào quá trình tính toán. Việc này không những giải quyết vấn đề cụ thể mà còn mang lại sự linh hoạt cho hệ thống, làm cho nó có khả năng ứng dụng rộng rãi hơn trong các điều kiện khác nhau. Hình 3.1. biểu thị hệ tọa độ và kiểu máy ảnh.



Hình 3.1. Hệ thống máy ảnh

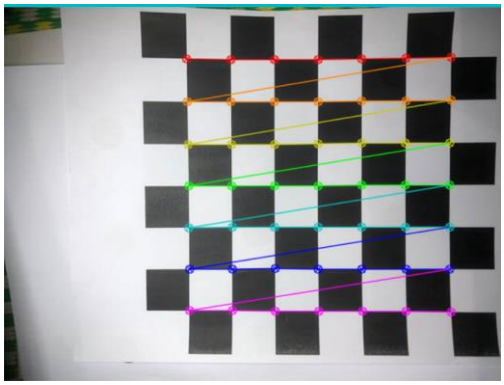
Để làm cho nó hoạt động, tôi đã thêm hệ số tỷ lệ s và ma trận camera A để đi đến u, v mà hiện đã được kích hoạt để giải tìm XYZ theo cách sau:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A[R \mid t]\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\left(s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} A^{-1}[\mathrm{R}|\mathrm{t}]^{-1} \right) = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Quá trình hiệu chỉnh máy ảnh là một bước quan trọng để đảm bảo rằng hình ảnh thu được từ camera là chính xác và đáng tin cậy. Để thực hiện việc này, dữ liệu đầu vào quan trọng nhất là một tập hợp các điểm trong thế giới thực 3D và các điểm hình ảnh 2D tương ứng của chúng. Đối với việc này, tôi sử dụng một bàn cờ có kích thước các ô là 2.5 cm, với mục tiêu đặt trước camera ở các tư thế khác nhau để thu được các góc quan sát đa dạng. Để tìm mẫu trong bàn cờ, tôi sử dụng hàm cv2.findChessboardCorners(). Nó trả về các điểm góc và giá trị trả về sẽ là True nếu lấy được mẫu. Các góc này sẽ được sắp xếp theo thứ tự (từ trái sang phải, từ trên xuống dưới). Hình ảnh mẫu được vẽ được ví dụ ở Hình 3.2. Bây giờ chúng ta đã có các điểm đối tượng và điểm hình ảnh, chúng ta đã sẵn sàng để hiệu chỉnh. Để làm được điều đó, tôi sử dụng hàm cv2.calibrateCamera(). Nó trả về ma trận camera, hệ số biến dạng, vectơ xoay và dịch chuyển, v.v.



Hình 3.2. Hình ảnh mẫu được vẽ

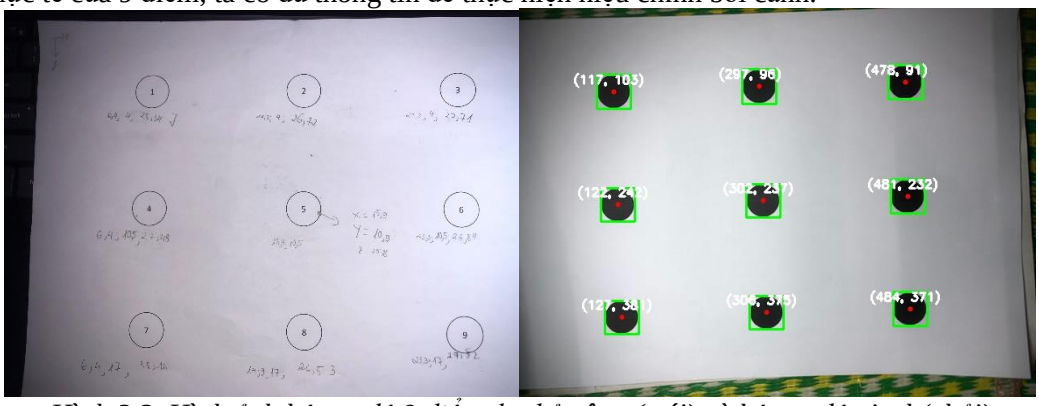
Sau khi hiệu chỉnh, ta thu được ma trận camera:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.3603e+03 & 0 & 2.2174e+02 \\ 0 & 1.3379e+03 & 5.2870e+02 \\ 0 & 0 & 1 \end{bmatrix}$$

B. Hiệu chỉnh bối cảnh

Trong đề tài này, tôi sử dụng một camera cố định duy nhất, điều đó có nghĩa là sau khi tôi hiệu chỉnh phối cảnh, mô hình sẽ bắt đầu hoạt động. Bước đầu tiên là xác định các giá trị Cx, Cy và z cho máy ảnh và tôi sử dụng Ma trận máy ảnh mới để tìm ra Cx=222 và Cy=529. Cx và Cy tương đương với các giá trị pixel u và v. Sau đó, tôi xác định vị trí điểm u = 222 và v = 342 theo cách thủ công. Thực hiên lặp đi lặp lại với

9 điểm bất kỳ, z chỉ đúng với điểm trung tâm và các điểm còn lại dùng lượng giác để tìm ra. Từ tọa độ pixel và tọa độ thực tế của 9 điểm, ta có đủ thông tin để thực hiện hiệu chỉnh bối cảnh.



Hình 3.3. Hình ảnh hệ tọa độ 9 điểm đo thủ công (trái) và hệ tọa độ pixel (phải)

Sau khi thực hiện hiệu chỉnh ta thu được hệ số tỷ lệ s = 957 và ma trận ngoại tại:

$$R|t = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} = \begin{bmatrix} 6.0134e-01 & -5.8232e-01 & -5.4708e-01 & 1.6654e+01 \\ 6.8973e-01 & 3.2689e-02 & 7.2333e-01 & 5.8645e+01 \\ -4.0333e-01 & -8.1230e-01 & 4.2130e-01 & -9.5277e+02 \end{bmatrix}$$

Sau khi tìm được các ma trận hiệu chỉnh, ta có thể dễ dàng tìm được tọa độ của độ vật trong môi trường thực khi biết tọa độ pixel của nó trong ảnh

C. Áp dụng xác định tọa độ 2D và 3D

- Xác định tọa độ 2D

Quá trình phát hiện vật thể qua phương pháp tìm contour là một giai đoạn quan trọng trong xử lý hình ảnh, đặc biệt là trong ngữ cảnh của việc theo dõi và nhận diện đối tượng. Trước hết, camera được khởi động để chụp ảnh background, và sau đó, thông qua quá trình loại bỏ nền, chúng ta tập trung vào đối tượng cần phát hiện. Áp dụng các kỹ thuật như làm mờ và thresholding giúp tăng cường sự nổi bật của đối tượng trong ảnh, giảm nhiễu và chuẩn bị cho bước tiếp theo.

Qua việc sử dụng hàm tìm contour của OpenCV, các đường biên của đối tượng được xác định, và chúng được sắp xếp theo độ lớn. Bằng cách vẽ bounding box xung quanh contour lớn nhất, chúng ta có thể thấy rõ đối tượng trong không gian hình ảnh. Điều này không chỉ giúp xác định hình dạng của đối tượng mà còn cung cấp thông tin quan trọng về vị trí của nó.

Cuối cùng, tọa độ tâm của vật thể được tính toán từ bounding box và sau đó được chuyển đổi từ tọa độ pixel trong ảnh sang tọa độ môi trường thực, đảm bảo tính chính xác và ứng dụng linh hoạt trong các ứng dụng thực tế. Tổng cộng, quá trình này không chỉ hỗ trợ trong việc phát hiện đối tượng mà còn mang lại thông tin quan trọng về vị trí và hình dạng của đối tượng trong không gian thực.

$$\left(s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} A^{-1}[\mathrm{R|t}]^{-1} \right) = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Xác định tọa độ 3D

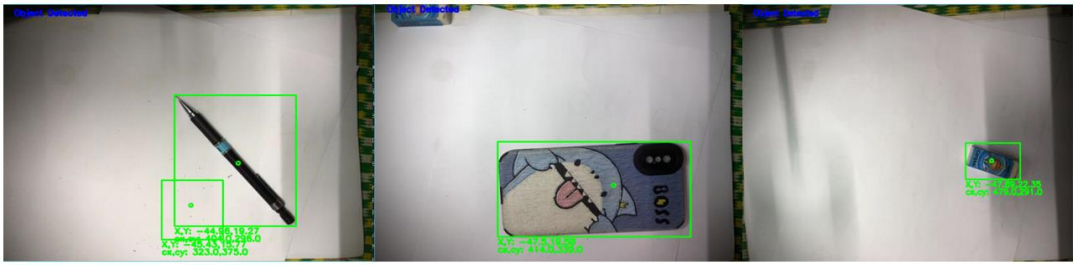
Trong quá trình xác định khối hộp bao quanh vật thể và trích xuất 8 đỉnh của khối hộp, chúng tôi đã áp dụng mô hình MediaPipe của Google, một công cụ mạnh mẽ với khả năng nhanh chóng và đa nền tảng. Mục tiêu chính của hệ thống là phát hiện và theo dõi các đối tượng 3D thông qua ảnh RGB đơn.

Đầu tiên, tôi tiến hành việc thu thập và gán nhãn dữ liệu, sử dụng dữ liệu Augmented Reality (AR) từ thiết bị di động và công cụ gán nhãn để tạo bounding box 3D cho các đối tượng trong không gian thực. Thông tin về tọa độ và hình dạng của đối tượng được cung cấp trong dữ liệu này, tạo nền tảng cho quá trình huấn luyện mô hình. Tiếp tục bằng việc tạo dữ liệu tổng hợp AR, đặt các đối tượng ảo vào cảnh thực để cải thiện độ chính xác và đa dạng của dữ liệu huấn luyện. Điều này giúp mô hình hiểu rõ về các biến thể của đối tượng và cách chúng xuất hiện trong không gian thực. Sau đó, huấn luyện mô hình để nó có khả năng

phát hiện và theo dõi các đối tượng 3D. Mô hình được điều chỉnh để đồng thời xác định vị trí và hình dạng của đối tượng trong không gian 3D, tận dụng thông tin chi tiết từ dữ liệu đã được gán nhãn và tổng hợp. Cuối cùng, khi mô hình đã được huấn luyện, chúng tôi có thể chuyển đổi tọa độ tâm của vật thể từ pixel sang tọa độ môi trường thực, áp dụng công thức chuyển đổi để định vị đối tượng trong không gian thực một cách chính xác và linh hoạt. Tổng cộng, quy trình này kết hợp sức mạnh của mô hình MediaPipe và thông tin đa dạng từ dữ liệu AR, đảm bảo khả năng phát hiện và theo dõi đối tượng 3D hiệu quả và đáng tin cậy trong các tình huống thực tế.

IV. KẾT QUẢ VÀ ĐÁNH GIÁ

- **Xác định tâm vật 2D:** Dưới đây là Hình 4.1 biểu thị kết quả sau khi thực hiện các vật mẫu là bút, ốp điện thoại và cục tẩy.



Hình 4.1. Kết quả xác định tâm vật 3D

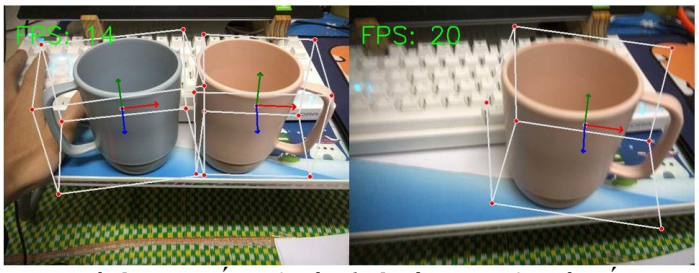
Kết quả của quá trình xác định tâm vật 2D trong không gian hình ảnh đã được đánh giá chặt chẽ để đảm bảo độ chính xác và tin cậy của hệ thống. Quá trình đánh giá này dựa trên so sánh tọa độ 2D dự đoán của hệ thống với tọa độ thực tế của các vật thể trong không gian hình ảnh. Kết quả cuối cùng cho thấy rằng hệ thống có khả năng xác định tâm vật 2D với mức sai số trung bình là 2mm.

Sai số này được đo lường bằng cách sử dụng dữ liệu từ một tập hợp các điểm được đặt và đo thủ công. Điều này tạo ra một bộ dữ liệu đánh giá chính xác để đối chiếu với kết quả dự đoán của hệ thống. Việc đánh giá này được thực hiện trong các điều kiện thực tế và đa dạng để đảm bảo khả năng ứng dụng rộng rãi của hệ thống.

Kết quả này là một bước quan trọng, chứng minh hiệu suất cao và độ chính xác của phương pháp xác định tâm vật 2D. Ứng dụng của hệ thống mở ra nhiều khả năng trong các lĩnh vực như theo dõi đối tượng, thực tế ảo, và giám sát trong thời gian thực.

- **Xác định tâm vật 3D:** Hệ thống đã đạt được thành công trong việc phát hiện và theo dõi đối tượng 3D trong không gian bằng cách sử dụng mô hình Mediapipe và kỹ thuật Bounding Box 3D. Kết quả cho thấy khả năng của hệ thống xác định tâm vật 3D với mức sai số trung bình là 2cm.

Mặc dù mức độ này có thể được chấp nhận trong nhiều tình huống, nhưng có thể ảnh hưởng đến độ chính xác của các ứng dụng đòi hỏi độ chính xác cao, đặc biệt là trong các lĩnh vực như y tế hoặc công nghiệp. Để cải thiện độ chính xác, có thể thực hiện điều chỉnh và tối ưu hóa hiệu suất của camera. Đồng thời, việc nâng cao độ chính xác trong quá trình hiệu chỉnh bối cảnh và thực hiện các bước xử lý dữ liệu có thể giúp giảm thiểu sai số và nâng cao hiệu suất. Mặc dù vẫn còn cơ hội để cải thiện, hệ thống vẫn là một công cụ mạnh mẽ với nhiều ứng dụng tiềm năng. Tuỳ thuộc vào yêu cầu cụ thể của ứng dụng, người sử dụng có thể điều chỉnh và tối ưu hóa hệ thống để đảm bảo độ chính xác phù hợp với mục tiêu cụ thể của ứng dụng.



Hình 4.2. Kết quả xác định tâm 3D của cái cốc

V. KẾT LUẬN

Trong nghiên cứu này, tôi đã tập trung vào việc phát triển một hệ thống xác định tọa độ tâm của vật thể từ hình ảnh thu được từ camera điện thoại iPhone X. Hệ thống này không chỉ giải quyết vấn đề trong không gian 2D mà còn mở rộng sang không gian 3D, sử dụng mô hình MediaPipe và kỹ thuật bounding box 3D.

Quá trình hiệu chỉnh máy ảnh và tái tạo 3D đã được thực hiện một cách cân nhắc và chi tiết. Bằng cách sử dụng các kỹ thuật xử lý hình ảnh và mô hình MediaPipe, hệ thống có khả năng xác định vị trí và hình dạng của đối tượng trong không gian thực. Kết quả cho thấy mức độ chính xác và độ ổn định đáng chú ý, đặc biệt là trong việc xác định tâm vật 2D với sai số trung bình là 2mm và tâm vật 3D với sai số trung bình là 2cm.

Tuy nhiên, cần lưu ý đến mức sai số và đề xuất các biện pháp cải thiện nhằm đáp ứng đòi hỏi của các ứng dụng đòi hỏi độ chính xác cao, như trong lĩnh vực y tế và công nghiệp. Việc thực hiện điều chỉnh hiệu suất camera, tối ưu hóa quá trình hiệu chỉnh bối cảnh, và cải thiện xử lý dữ liệu là những hướng đi có thể được thực hiện để giảm thiểu sai số.

Mặc dù vẫn còn cơ hội để cải thiện, hệ thống vẫn là một công cụ mạnh mẽ với nhiều ứng dụng tiềm năng trong theo dõi, giám sát, thực tế ảo, và nhiều lĩnh vực khác. Tùy thuộc vào yêu cầu cụ thể của ứng dụng, người sử dụng có thể điều chỉnh và tối ưu hóa hệ thống để đảm bảo độ chính xác phù hợp với mục tiêu cụ thể của họ.

VI. TÀI LIỆU THAM KHẢO

[1] S. Bharathi, P.K. Pareek, B.R.S. Rani, D.R. Chaitra, "3-Dimensional Object Detection Using Deep Learning Techniques," in N.R. Shetty, N.H. Prasad, N. Nalini (eds), Advances in Computing and Information, ERCICA 2023, vol. 1104, Lecture Notes in Electrical Engineering. Springer, Singapore, 2024.

[2] FDX Labs. "Calculate X, Y, Z Real-World Coordinates from a Single Camera Using OpenCV." FDX Labs, https://www.fdxlabs.com/calculate-x-y-z-real-world-coordinates-from-a-single-camera-using-opencv/

[3] OpenCV Documentation. "Camera Calibration and 3D Reconstruction." OpenCV Documentation, https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html.

[4] OpenCV Documentation. "Camera Calibration using OpenCV - Python." OpenCV Documentation, https://docs.opencv.org/3.3.0/dc/dbb/tutorial_py_calibration.html.

[5] Kaustubh Sadekar, Satya Mallick, "Camera Calibration using OpenCV." LearnOpenCV, https://learnopencv.com/camera-calibration-using-opencv/.