

PHÁT TRIỂN HỆ THỐNG THỊ GIÁC ROBOT CẤP PHÔI TỰ ĐỘNG CHO MÁY TIỆN CNC

DEVELOPMENT OF ROBOTIC VISION SYSTEM FOR AUTOMATED WORKPIECE FEEDING IN CNC LATHES

Vũ Viết Toàn, Lê Như Kiệt, Nguyễn Thành Hùng*

Đại học Bách khoa Hà Nội

*Email: hung.nguyenthanh@hust.edu.vn

TÓM TẮT

Bài báo này trình bày một hệ thống thị giác robot được phát triển nhằm tự động cấp phôi cho máy tiện CNC. Hệ thống phần cứng bao gồm camera 3D RealSense gắn trên cánh tay robot Igus với 5 bậc tự do. Phần mềm điều khiển được xây dựng dựa trên công nghệ xử lý ảnh 3 chiều kết hợp học sâu. Camera 3D thu nhận hình ảnh các đối tượng dưới dạng đám mây điểm, từ đó thuật toán thị giác 3D được phát triển để xác định chính xác vị trí và hướng của vật thể. Các thông tin này được sử dụng để điều khiển robot thực hiện thao tác gấp và đặt vật thể vào vị trí xác định trước. Tính khả thi và hiệu quả của hệ thống đã được xác minh thông qua thực nghiệm. Kết quả cho thấy hệ thống có thể gấp và đặt chính xác các vật thể được sắp xếp ngẫu nhiên trong không gian làm việc của camera 3D.

Từ khóa: Robot gấp vật; Xử lý ảnh 3 chiều; Công nghệ học sâu; Định vị đối tượng; Cấp phôi tự động.

ABSTRACT

This paper presents a robotic vision system for automatic workpiece feeding on a CNC lathe. The hardware setup includes a 3D RealSense camera mounted on a five-degree-of-freedom Igus robotic arm. The control software is developed using 3D image processing and deep learning technologies. The 3D camera captures point cloud data of objects, and a 3D vision algorithm is employed to determine their position and orientation. This information is used to guide the robotic arm to accurately pick and place the objects at predefined locations. The feasibility and effectiveness of the developed system are validated through experiments. Results show that the system can accurately handle randomly positioned objects within the camera's workspace.

Keywords: Robotic bin picking; 3D image processing; Deep learning; Object localization; Automated workpiece feeding. 

1. ĐẶT VẤN ĐỀ

Trong ngành công nghiệp hiện đại, cánh tay robot được ứng dụng rộng rãi để thay thế con người trong các công việc như đóng gói, cấp phối, gia công và lắp ráp. Trước nhu cầu xử lý các sản phẩm có hình dạng phức tạp ngày càng tăng, robot cần có độ linh hoạt và đa năng cao để đáp ứng sản xuất nhanh, hiệu quả và dễ dàng tùy chỉnh cho nhiều loại sản phẩm. Một trong những giải pháp hiệu quả là tích hợp hệ thống thị giác máy tính vào cánh tay robot [1].

Tuy nhiên, việc xác định vị trí và hướng của vật thể trong không gian 3D vẫn là một thách thức lớn. Nguyên nhân chủ yếu do dữ liệu đầu vào từ camera chỉ cung cấp thông tin không đầy đủ hoặc bị che khuất. Ngoài ra, các vật thể có thể được đặt ngẫu nhiên, thậm chí tiếp xúc hoặc chồng lên nhau, làm phức tạp quá trình nhận dạng. Để giải quyết vấn đề này, các hệ thống thị giác 3D đã được tích hợp nhằm hỗ trợ nhận dạng và định vị vật thể nhanh chóng và hiệu quả, phục vụ quá trình gấp – đặt bằng robot [2-7].

Các hệ thống thị giác 3D truyền thống thường sử dụng camera 3D chuyên dụng như camera ánh sáng cấu trúc hoặc ToF (Time-of-Flight), được gắn trực tiếp trên robot. Những camera này giúp thu thập thông tin chiều sâu, từ đó xác định chính xác vị trí, hướng và hình dạng của vật thể, kể cả khi được sắp xếp ngẫu nhiên. So với hệ thống thị giác 2D, giải pháp này mang lại hiệu quả vượt trội trong các môi trường phức tạp [2].

Hiện nay, công nghệ thị giác 3D tích hợp AI đang tạo ra bước đột phá lớn, giúp robot tương tác tinh vi hơn với môi trường. Bằng cách kết hợp dữ liệu độ sâu với các thuật toán học

máy tiên tiến, robot có thể nhận dạng và thao tác chính xác với các vật thể trong môi trường không cấu trúc, như thùng chứa. Công nghệ này làm tăng đáng kể hiệu suất và độ chính xác của các hệ thống tự động hóa trong sản xuất hiện đại [3-7].

Mặc dù cả thị giác 3D truyền thống và thị giác 3D tích hợp AI đều có thể được ứng dụng trong bài toán gấp – đặt, điểm khác biệt lớn nhất nằm ở khả năng xử lý môi trường phức tạp và thay đổi liên tục. Công nghệ truyền thống gặp khó khăn khi đối mặt với vật thể có vị trí, hướng không cố định, trong khi công nghệ dựa trên AI lại có khả năng học và thích ứng tốt, mang lại tính linh hoạt cao hơn.

Bài báo này đề xuất một hệ thống kết hợp xử lý ảnh 3D với công nghệ thị giác máy tính dựa trên AI để giải quyết bài toán cấp phối tự động cho máy tiện CNC. Bố cục bài báo như sau: Phần 2 trình bày tổng quan hệ thống. Phần 3 mô tả phương pháp nhận dạng và định vị vật thể. Phần 4 giới thiệu các thí nghiệm và kết quả. Cuối cùng, phần 5 là kết luận và định hướng phát triển tiếp theo.

2. TỔNG QUAN HỆ THỐNG

Hình 1 minh họa hệ thống robot cấp phối tự động được xây dựng trong nghiên cứu này. Hệ thống sử dụng cánh tay robot Igus Robolink DCI-5S-M với 5 bậc tự do, có tầm với 680 mm, độ chính xác 0,5 mm và khả năng tải 0,5 kg. Camera 3D Intel RealSense D435i được gắn cố định trên robot, hỗ trợ thu nhận hình ảnh với độ phân giải RGB 1920×1080 ở 30 FPS và độ phân giải chiều sâu 1280×720 ở 30 FPS, trong phạm vi đo từ 0,2 m đến 10 m.

Vị trí và hướng của vật thể được xác định thông qua thuật toán nhận dạng và định vị

3D dựa trên công nghệ trí tuệ nhân tạo. Sau đó, bộ kẹp gắn trên tay robot sẽ được điều khiển để gấp và đặt vật thể vào vị trí xác định trước.



Hình 1. Hệ thống thị giác robot cấp phôi tự động cho máy tiện CNC.

3. PHƯƠNG PHÁP NHẬN DIỆN VÀ XÁC ĐỊNH VỊ TRÍ CỦA VẬT

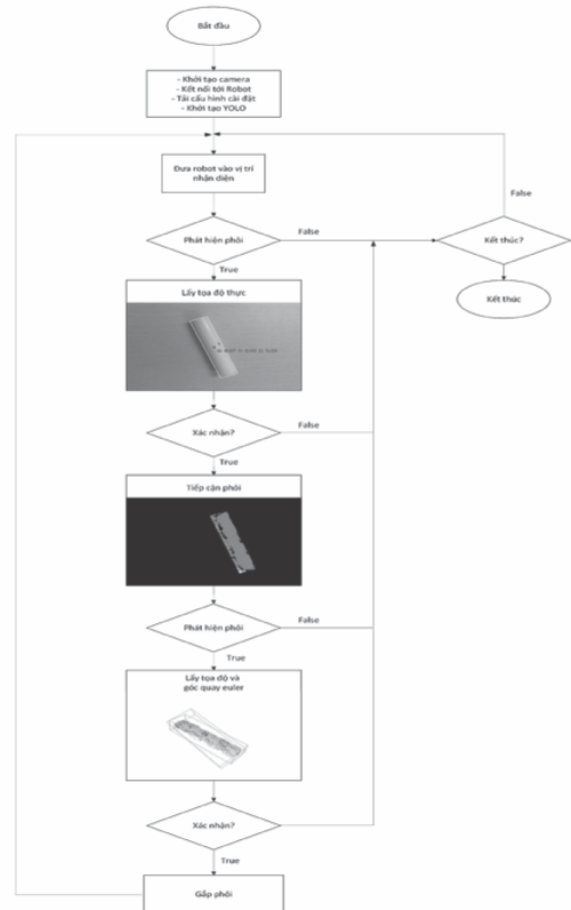
Hình 2 mô tả quy trình gấp và đặt phôi sử dụng hệ thống thị giác cấp phôi tự động được phát triển. Khi hệ thống khởi động, phần mềm sẽ tiến hành khởi tạo camera 3D, kết nối với robot, tải các cấu hình cài đặt và khởi động thuật toán học sâu YOLOv11 [8]. Sau khi hoàn tất, robot sẽ di chuyển đến vùng làm việc để nhận diện các phôi.

Dữ liệu đầu ra từ camera 3D bao gồm thông tin màu sắc (RGB) và dữ liệu độ sâu. Quá trình nhận dạng và định vị vị trí phôi diễn ra theo hai bước:

- Bước 1: Xác định vị trí ban đầu của phôi. Ảnh RGB được xử lý bằng thuật toán YOLOv11-OBb để phát hiện vùng chứa phôi. Sau đó, vùng ảnh tương ứng được kết hợp với ảnh độ sâu để tạo đám mây điểm của phôi. Trọng tâm của đám mây điểm được sử dụng làm vị trí gấp.

- Bước 2: Xác định chính xác vị trí và

hướng gấp. Robot được điều khiển tiếp cận gần hơn với phôi dựa trên dữ liệu thu được ở bước 1. Quá trình trích xuất đám mây điểm được lặp lại để nâng cao độ chính xác. Hướng gấp được tính toán dựa trên các trục chính của hình hộp bao định hướng của đám mây điểm.



Hình 2. Thuật toán điều khiển robot gấp và đặt phôi tiện.

3.1. Phát hiện phôi sử dụng YOLOv11-OBb

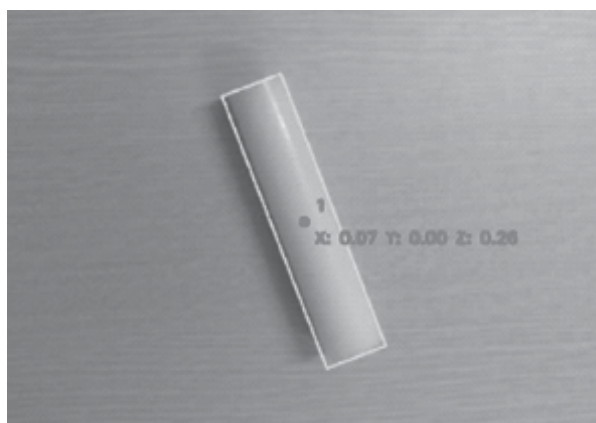
YOLOv11 là phiên bản mới nhất trong chuỗi các mô hình YOLO (You Only Look Once) do Ultralytics phát triển. Mô hình này cải tiến hiệu suất và độ chính xác trong nhận dạng vật thể, đồng thời tối ưu tài nguyên tính toán so với các phiên bản trước.



YOLOv11-OBb (Oriented Bounding Box) là biến thể đặc biệt được thiết kế để phát hiện các vật thể không thẳng hàng với trục hình ảnh – đặc biệt hữu ích trong các ứng dụng như ảnh vệ tinh, ảnh công nghiệp. Mô hình sử dụng định dạng bao vật thể bằng 4 điểm góc, cho phép nhận diện chính xác hình dạng và hướng của vật.

Mô hình này đặc biệt hiệu quả trong việc xử lý các nhiệm vụ nhận diện vật thể phức tạp trong môi trường mà các vật thể như bể chứa, đường ống và tàu được quan sát từ trên cao, làm cho nó lý tưởng cho các ngành như quy hoạch đô thị, giám sát cơ sở hạ tầng, và quản lý cảng biển. Khả năng xử lý và nhận diện vật thể từ các góc nhìn khác nhau cho phép đạt được kết quả nhận diện chính xác và đáng tin cậy hơn, điều này rất quan trọng cho việc giám sát và quản lý hiệu quả.

YOLOv11-OBb có cấu trúc bao gồm các lớp tích chập và các khối chuyên biệt như C3k2 và SPPF (Spatial Pyramid Pooling – Fast), được tối ưu hóa để tập trung vào các vùng quan trọng trong ảnh. Điều này giúp mô hình không chỉ chính xác mà còn hiệu quả về mặt tính toán. Kết quả nhận diện phôi sử dụng YOLOv11-OBb được thể hiện trong hình 3.



Hình 3. Kết quả nhận diện phôi sử dụng YOLOv11-OBb.

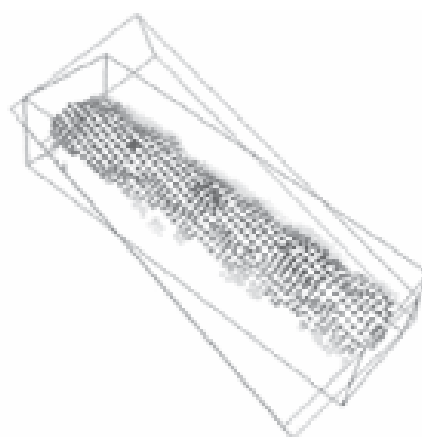
3.2. Trích xuất đám mây điểm của phôi

Vùng ảnh phôi phát hiện trong ảnh RGB sẽ được dùng làm mặt nạ để trích xuất đám mây điểm tương ứng từ ảnh độ sâu. Sau đó, đám mây điểm được lấy mẫu nhằm giảm số lượng điểm, đồng thời lọc nhiễu bằng phương pháp thống kê [9] sử dụng thư viện Open3D [10].

Tiếp theo, một hình hộp bao định hướng được tính toán từ đám mây điểm. Các trục chính của hình hộp này đại diện cho các hướng gấp vật (hình 4). Vị trí gấp vật được xác định là trọng tâm của đám mây điểm.



(a)



(b)

Hình 4. (a) Đám mây điểm của phôi; (b) Hình hộp bao định hướng của đám mây điểm.

3.3. Hiệu chuẩn tọa độ robot và camera

Hiệu chuẩn “eye-in-hand” là kỹ thuật thường dùng trong các hệ thống thị giác gắn trên tay robot. Mục tiêu là xác định chính xác quan hệ hình học giữa hệ tọa độ của camera và hệ tọa độ của tay robot.

Các bước hiệu chuẩn như sau:

- Bước 1: Đặt một bảng checkerboard cố định trong không gian; điều khiển robot để camera quan sát toàn bộ bảng.

- Bước 2: Ghi nhận ảnh và tọa độ góc Euler của robot tại mỗi lần chụp.

- Bước 3: Dùng OpenCV [11] để tính toán vị trí checkerboard trong hệ tọa độ camera,

thu được tập ma trận tương ứng: $T_{\text{camera}}^{\text{checkerboard}}$.

- Bước 4: Từ tọa độ và góc quay của tay robot, tính toán tập ma trận tương ứng: $T_{\text{base}}^{\text{end-effector}}$.

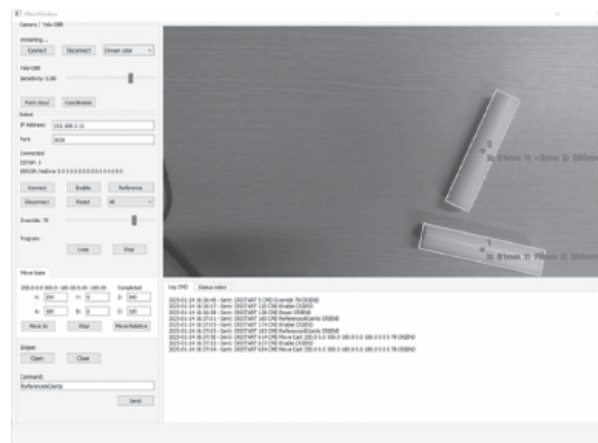
- Bước 5: Dùng công cụ hiệu chuẩn trong OpenCV để tính ra ma trận chuyển đổi giữa hai hệ tọa độ.

3.4. Điều khiển robot

Hệ thống sử dụng giao thức CRI (Common Robot Interface) để gửi lệnh điều khiển robot. Đây là một chuẩn giao tiếp hỗ trợ tích hợp các robot công nghiệp khác nhau với hệ thống điều khiển.

CRI giúp các kỹ sư phát triển các ứng dụng điều khiển và giám sát robot một cách linh hoạt, nhanh chóng và hiệu quả hơn. Giao diện điều khiển tích hợp giữa robot và camera được thể hiện trong hình dưới.

Giao diện điều khiển robot và camera được thể hiện trong hình 5.



Hình 5. Giao diện điều khiển robot và camera.

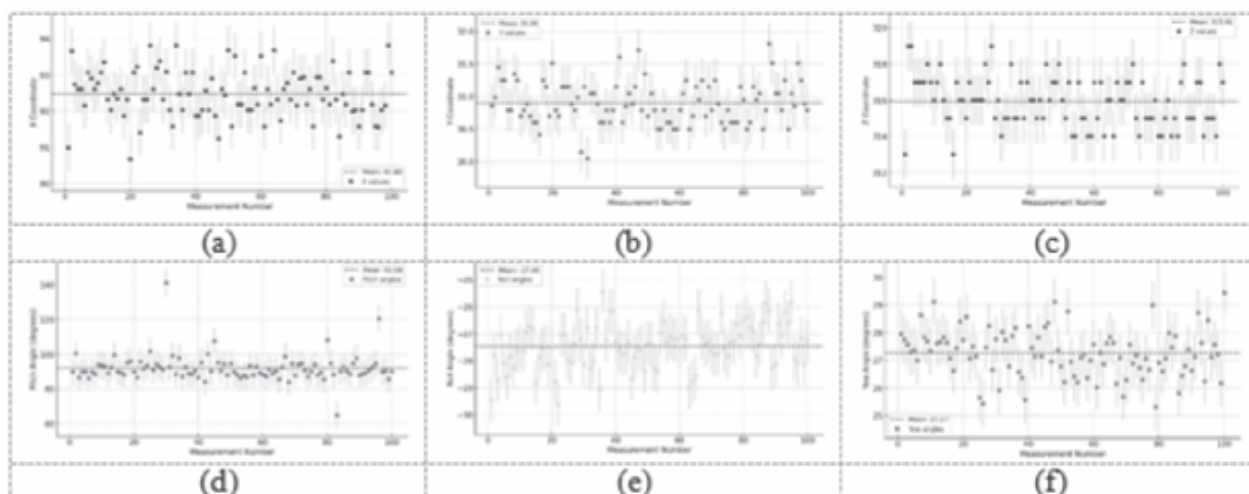
4. KẾT QUẢ THỬ NGHIỆM VÀ BÀN LUẬN

Quy trình thực nghiệm bao gồm các bước: chuẩn bị hệ thống, thu thập dữ liệu, nhận dạng và hiệu chuẩn. Hệ thống được kết nối và cài đặt các phần mềm CPRog và OpenCV để lập trình và xử lý dữ liệu. Camera 3D tiến hành quét bề mặt vùng làm việc, thu thập ảnh RGB và dữ liệu độ sâu. Sau đó, thuật toán YOLOv11 được sử dụng để nhận dạng phôi, trong khi thư viện Open3D xử lý đám mây điểm và xác định vị trí của phôi trong không gian.

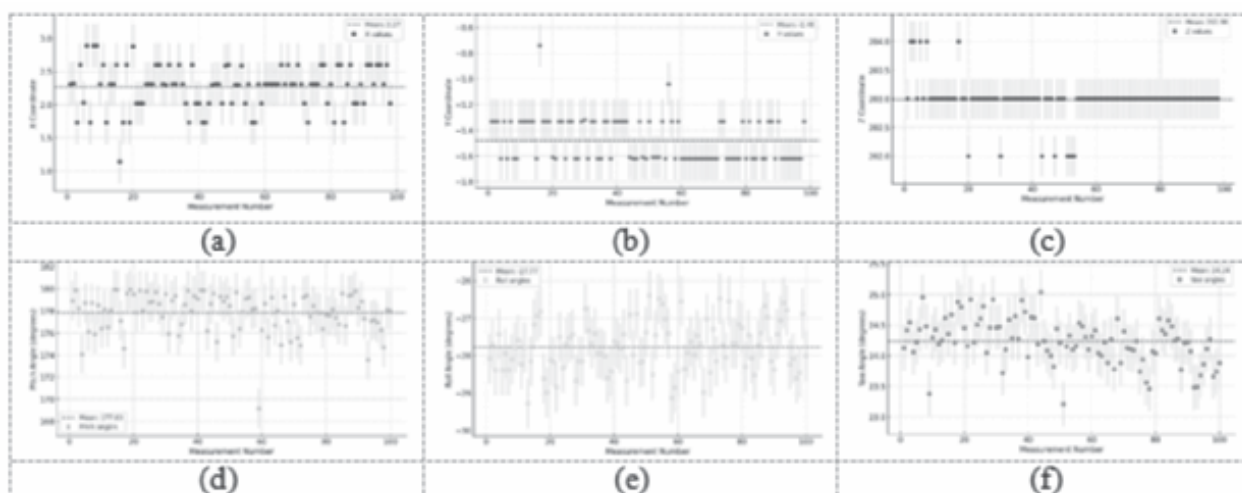
Các dữ liệu tọa độ sau khi xử lý được chuyển đổi sang hệ tọa độ của robot thông qua bước hiệu chuẩn. Quá trình này giúp nâng cao độ chính xác và độ ổn định của hệ thống. Phôi sử dụng trong thí nghiệm được làm từ nhựa POM, có kích thước đường kính 2 cm và chiều dài 10 cm. Tổng số lần thực nghiệm là 100.

Hình 6 thể hiện kết quả xác định vị trí và hướng của phôi trong trường hợp chỉ sử dụng “Bước 1 – Xác định vị trí ban đầu”. Hình 7 minh họa kết quả tương ứng trong trường hợp

sử dụng cả “Bước 1 – Xác định vị trí ban đầu” và “Bước 2 – Xác định chính xác vị trí và hướng gấp”. Độ lệch chuẩn về vị trí và góc định hướng (góc Euler) trong hai trường hợp này được tổng hợp trong Bảng 1.



Hình 6. Kết quả xác định vị trí: x – (a), y – (b), z – (c) và hướng của phôi: Pitch – (d), Roll – (e), Yaw – (f) trong trường hợp chỉ sử dụng Bước 1.



Hình 7. Kết quả xác định vị trí: x – (a), y – (b), z – (c) và hướng của phôi: Pitch – (d), Roll – (e), Yaw – (f) trong trường hợp sử dụng cả Bước 1 và Bước 2.

Bảng 1. Độ lệch chuẩn về vị trí và hướng gấp phôi.

Bước thực hiện	Độ lệch chuẩn					
	x (mm)	y (mm)	z (mm)	Pitch (rad)	Roll (rad)	Yaw (rad)
Bước 1	0,65	0,32	1,37	0,1342	0,0141	0,0148
Bước 1 + Bước 2	0,33	0,17	0,35	0,0311	0,0117	0,0065

Kết quả thực nghiệm cho thấy rằng việc sử dụng thuật toán YOLOv11-OBb trong giai đoạn đầu giúp phát hiện nhanh vùng phôi trong ảnh RGB, tuy nhiên độ chính xác vẫn còn hạn chế do ảnh hưởng của độ phân giải chiều sâu, sự nhiễu từ môi trường hoặc độ tương phản thấp giữa phôi và nền. Điều này lý giải vì sao độ lệch chuẩn ở bước 1 còn tương đối lớn, đặc biệt theo trục z và góc Pitch. Việc tiếp cận gần hơn phôi trong bước 2 không chỉ cải thiện chất lượng ảnh độ sâu mà còn giúp tăng mật độ điểm trong đám mây điểm, từ đó tăng độ chính xác khi trích xuất hình hộp bao định hướng. Nhờ vậy, sai số về vị trí và hướng gấp đã được giảm đáng kể, đặc biệt trong các thông số Pitch và Yaw – vốn ảnh hưởng lớn đến thao tác gấp chính xác của robot.

Điều này nhấn mạnh vai trò của việc kết hợp thị giác với thuật toán di chuyển thông minh của robot trong các ứng dụng thực tế: thay vì chỉ dựa vào một lần nhận diện từ xa, hệ thống được thiết kế để làm giàu thông tin qua tiếp cận gần, góp phần cải thiện hiệu suất tổng thể. Ngoài ra, hệ thống thể hiện tính linh hoạt tốt khi làm việc với các vật thể được đặt ngẫu nhiên trong vùng làm việc – điều mà các giải pháp thị giác truyền thống gặp nhiều khó khăn. Đây là tiền đề quan trọng cho việc áp dụng hệ thống vào các dây chuyền sản xuất có tính biến động cao, sản xuất theo lô nhỏ hoặc sản phẩm tùy chỉnh.

Tuy nhiên, vẫn tồn tại một số hạn chế cần được khắc phục trong tương lai, bao gồm:

- Tăng cường tính ổn định của việc hiệu chuẩn giữa hệ tọa độ camera và robot trong các điều kiện ánh sáng thay đổi.

- Rút ngắn thời gian xử lý bằng cách tối ưu thuật toán và tận dụng GPU hiệu quả hơn.

- Khả năng phát hiện các vật thể có hình dạng phức tạp hoặc bị che khuất một phần.

Những cải tiến này sẽ giúp hệ thống hoàn thiện hơn và dễ dàng ứng dụng trong các môi trường công nghiệp thực tế với yêu cầu cao về hiệu suất và độ tin cậy.

5. KẾT LUẬN

Bài báo đã đề xuất và phát triển một hệ thống cấp phôi tự động cho máy tiện CNC, sử dụng cánh tay robot 5 bậc tự do tích hợp camera 3D và các thuật toán học sâu. Hệ thống áp dụng mô hình YOLOv11 để nhận dạng phôi từ ảnh RGB, đồng thời sử dụng thư viện Open3D để xử lý đám mây điểm và định vị chính xác vật thể trong không gian 3D. Ngoài ra, kỹ thuật hiệu chuẩn “eye-in-hand” giúp đồng bộ hệ tọa độ giữa camera và tay robot, cho phép điều khiển chính xác các thao tác gấp – đặt. Kết quả thực nghiệm cho thấy hệ thống hoạt động ổn định và đạt độ chính xác cao, đặc biệt sau khi thực hiện bước tiếp cận gần vật thể để tinh chỉnh vị trí và hướng gấp. Hệ thống được phát triển không chỉ góp phần nâng cao hiệu quả tự động hóa trong gia công cơ khí mà còn có tiềm năng mở rộng sang các ứng dụng sản xuất thông minh khác trong tương lai.

Lời cảm ơn:

Bài báo được thực hiện trong khuôn khổ đề tài: “Nghiên cứu và phát triển hệ thống điều khiển robot cộng tác tích hợp thị giác máy tính và trí tuệ nhân tạo”, mã số B2024.BKA.09. Nhóm nghiên cứu xin trân trọng cảm ơn Bộ Giáo dục và Đào tạo đã tài trợ cho nghiên cứu này. ❖

Ngày nhận bài: **20/3/2025**

Ngày phản biện: **02/4/2025**



Tài liệu tham khảo:

- [1]. B. van Eden and B. Rosman, “*An overview of robot vision*”. 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Bloemfontein, South Africa, 2019, pp. 98-104.
- [2]. Nguyen, T.-H., Chen, L.-C., & Dang, T.-V. (2016), “*Automated Robotic Bin Picking System of Randomly Stacked Objects*”. National Conference on Mechanical & Transportation Engineering, Hanoi, Vietnam.
- [3]. Lee, S., Islam, N. U., & Lee, S. (2020), “*Robust image completion and masking with application to robotic bin picking*”. Robotics and Autonomous Systems, 131(103563), 103563. <https://doi.org/10.1016/j.robot.2020.103563>
- [4]. Li, X., Cao, R., Feng, Y., Chen, K., Yang, B., Fu, C.-W., Li, Y., Dou, Q., Liu, Y.-H., & Heng, P.-A. (2022), “*A Sim-to-real object recognition and localization framework for industrial robotic bin picking*”. IEEE Robotics and Automation Letters, 7(2), 3961-3968. <https://doi.org/10.1109/lra.2022.3149026>
- [5]. Tsai, C.-H., Hernandez, E. E., You, X.-W., Lin, H.-Y., & Chang, J.-Y. (2023), “*RoboTwin metaverse platform for robotic random bin picking*”. Applied Sciences (Basel, Switzerland), 13(15), 8779. <https://doi.org/10.3390/app13158779>
- [6]. Sun, H., Zhang, Z., Wang, H., Wang, Y., & Cao, Q. (2024), “*A novel robotic grasp detection framework using low-cost RGB-D camera for industrial bin picking*”. IEEE Transactions on Instrumentation and Measurement, 73, 1-12. <https://doi.org/10.1109/tim.2023.3346531>
- [7]. Dolezel, P., Stursa, D., & Kopecky, D. (2024), “*Memory efficient deep learning-based grasping point detection of nontrivial objects for robotic bin picking*”. Journal of Intelligent & Robotic Systems, 110(3). <https://doi.org/10.1007/s10846-024-02153-9>
- [8]. Glenn Jocher and Jing Qiu (2024), “*Ultralytics YOLO11*”.
- [9]. Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M., & Beetz, M. (2008), “*Towards 3D Point cloud based object maps for household environments*”. Robotics and Autonomous Systems, 56(11), 927-941.
- [10]. Zhou, Q.-Y., Park, J., & Koltun, V. (2018), “*Open3D: A modern library for 3D data processing*”. arXiv:1801.09847.
- [11]. Bradski, G. (2000), “*The OpenCV Library*”. Dr. Dobb's Journal of Software Tools.