

# ROBOT NHỎ CỎ TỰ ĐỘNG DỰA TRÊN PHÂN TÍCH ẢNH SỬ DỤNG MÔ HÌNH HỌC SÂU

## AN AUTOMATIC WEEDING ROBOT BASED ON IMAGE ANALYSIS USING DEEP LEARNING MODEL

Phan Trần Đăng Khoa<sup>1\*</sup>, Từ Phương Nguyên<sup>1</sup>, Đỗ Tuấn Sơn<sup>1</sup>, Nguyễn Quang Hiếu<sup>1</sup>, Vũ Văn Thanh<sup>1</sup>

<sup>1</sup>Trường Đại học Bách khoa - Đại học Đà Nẵng

\*Tác giả liên hệ: ptdkhoa@dut.udn.vn

(Nhận bài: 01/02/2021; Chấp nhận đăng: 15/3/2021)

**Tóm tắt** - Hiện nay, việc lạm dụng hóa chất diệt cỏ không chỉ gây tác động xấu đến môi trường và sức khỏe con người mà còn dẫn đến việc hình thành ngày càng nhanh các loại cỏ có khả năng kháng thuốc. Do đó, việc nghiên cứu các phương pháp diệt cỏ không sử dụng hóa chất đang ngày càng được quan tâm. Trong bài báo này, nhóm tác giả đề xuất một thiết kế robot nhỏ cỏ tự động dựa trên phân tích hình ảnh. Từ đó, chế tạo một nguyên mẫu để kiểm thử và đánh giá. Mô hình học sâu được áp dụng với cơ sở dữ liệu được thu thập thực tế để phát hiện các đối tượng cây trồng và cỏ dại. Nhóm tác giả đề xuất phương pháp để xác định tọa độ 3D của đối tượng quan tâm dựa trên mô hình camera. Kết quả thực nghiệm cho thấy, hệ thống có độ chính xác cao trong phát hiện, phân loại và xác định tọa độ 3D của các đối tượng quan tâm.

**Từ khóa** - Học sâu; Phát hiện đối tượng; Diệt cỏ tự động; Robot

### 1. Giới thiệu

Trong thế kỷ trước, sản xuất nông nghiệp đã đạt được các thành tựu vượt trội nhờ vào việc canh tác trên diện rộng kết hợp với cơ khí hóa các công cụ sản xuất. Ngoài ra, các giống tăng trưởng và biến đổi gen, cũng như thuốc trừ sâu, diệt cỏ và phân bón cũng đã góp phần nâng cao năng suất sản xuất. Tuy nhiên, xu hướng này cũng gây ra một số vấn đề. Việc sử dụng các hóa chất để trừ sâu và diệt cỏ trên diện rộng tác động xấu đến môi trường và dẫn đến tình trạng kháng thuốc ở các động thực vật liên quan. Theo thống kê ở báo cáo [1], tình trạng cỏ kháng thuốc đã làm tăng chi phí sản xuất lên 27% trên một mẫu Anh tại Australia do tăng chi phí diệt cỏ và thiệt hại về năng suất cây trồng. Tại Mỹ, các nông dân phải trả thêm 150 USD trên một mẫu Anh để nhổ cỏ thủ công khi các biện pháp diệt cỏ bằng hóa chất không còn hiệu quả do cỏ kháng thuốc. Tuy chưa có thống kê cụ thể tại Việt Nam nhưng một số nghiên cứu trong nước cũng đã đánh giá mức độ cỏ kháng thuốc và đưa ra các khuyến cáo [2].

Nhằm giải quyết vấn đề này, bên cạnh các biện pháp sử dụng thuốc diệt cỏ hợp lý thì việc áp dụng các kỹ thuật và công nghệ tiên tiến để thay thế một phần hoặc hoàn toàn phương pháp sử dụng hóa chất đang được quan tâm nghiên cứu. Gần đây, các robot diệt cỏ tự động với khả năng phân loại cây trồng và cỏ dại cho phép cắt giảm lượng hóa chất sử dụng vì không phải phun trên diện rộng [3-5]. Một số nghiên cứu khác đề xuất các thiết kế kết hợp giữa xe tự hành và máy bay không người lái để phát triển một nền tảng robot quản lý cây trồng và cỏ dại [6]. Đối với các mô hình này thì khối

**Abstract** - Currently, the application of herbicides not only has negative impacts on the environment and human health, but also leads to rapidly increasing herbicide resistance in weeds. Therefore, the study of weeding methods without using herbicides attracts interest of researchers. In this paper, we propose the design of an automatic weeding robot based on image analysis. Based on this design, we build a prototype for testing and evaluation. A deep learning model is applied on a dataset collected in real environment to detect crops and weeds. We also propose a method to determine the 3D coordinates of objects based on camera calibration. Experimental results show that the system has high accuracy for object detection and classification, and localization of objects in 3D environment.

**Key words** - Deep learning; Object detection; Automatic weeding; Robot

phát hiện và phân loại các đối tượng cây trồng, cỏ dại đóng vai trò quan trọng. Việc phát hiện cỏ dại thường được tiến hành thông qua hình ảnh thu hoặc từ camera màu thông thường hoặc kết hợp với camera cận hồng ngoại. Việc sử dụng camera thông thường có ưu điểm về giá thành. Tuy có giá thành cao nhưng camera cận hồng ngoại lại có lợi thế hơn do bức xạ trong dải tần này được phản xạ mạnh bởi lá cây. Phương pháp chung để phát hiện cây trồng, cỏ dại là phân vùng ảnh theo màu sắc trong các không gian màu khác nhau như HSV, Lab, ... [3]. Tuy nhiên, phương pháp này bị ảnh hưởng nhiều bởi sự đa dạng và phức tạp về màu sắc của cây trồng và nhiễu ánh sáng từ môi trường. Việc phân loại cây trồng và cỏ dại được thực hiện thông qua các mô hình học máy (machine learning) với các đặc trưng đầu vào liên quan đến hình dáng, cấu trúc, gradient, ... [4, 5].

Trong thời gian gần đây, với sự phát triển của kỹ thuật học sâu (deep learning), các bài toán phát hiện và phân loại đối tượng được giải quyết hiệu quả. Các nghiên cứu [7-12] đã đề xuất các mô hình phát hiện đối tượng, phù hợp với các tiêu chí ứng dụng khác nhau như độ chính xác, tốc độ xử lý, ... Các mô hình phát hiện đối tượng dựa trên kỹ thuật học sâu được chia thành 2 loại cơ bản như sau [7]: Thứ nhất, các mô hình như R-CNN [8], Fast R-CNN [9], ... đưa ra các vùng đề xuất và thực hiện phân loại đối tượng cho các vùng này; Thứ hai, các mô hình như YOLO [10, 11], SSD [12], ... thực hiện phép hồi quy để dự đoán các thông tin về vị trí và phân loại đối tượng dựa trên các đặc trưng phân tích được từ ảnh thông qua

<sup>1</sup> The University of Danang - University of Science and Technology (Phan Tran Dang Khoa, Tu Phuong Nguyen, Do Tuan Son, Nguyen Quang Hieu, Thanh Vu Van)

mạng nơ-ron tích chập. Các mô hình dựa trên vùng đề xuất có lợi thế về độ chính xác đối với định vị đối tượng nhưng có tốc độ xử lý chậm. Trong khi đó, các mô hình dựa trên hồi quy có ưu điểm về tốc độ xử lý nhanh và độ chính xác đối với phân loại đối tượng.

Trong bài báo này, nhóm tác giả đề xuất một thiết kế robot diệt cỏ tự động với khối phát hiện, phân loại và định vị đối tượng được xây dựng dựa trên mô hình học sâu. So với các nghiên cứu liên quan [3-6], nhóm tác giả có 3 đóng góp chính trong nghiên cứu này. Cụ thể như sau: Thứ nhất, nhóm tác giả đề xuất một thiết kế khung robot với cơ cấu đầu kẹp để nhỏ cỏ nhằm tránh sử dụng hóa chất diệt cỏ. Dựa vào thiết kế này, chế tạo một nguyên mẫu để kiểm thử và đánh giá; Thứ hai, áp dụng mô hình học sâu với cơ sở dữ liệu được thu thập thực tế để phát hiện và phân loại đối tượng đối tượng cây trồng, cỏ dại; Thứ ba, đề xuất phương pháp xác định tọa độ 3D của các điểm quan tâm từ tọa độ ảnh tương ứng dựa vào mô hình camera.

**2. Tổng quan về hệ thống**

Sơ đồ khối của hệ thống được mô tả ở Hình 1. Hình ảnh thu được từ camera sẽ sử dụng để phát hiện cỏ dại. Một mô hình học sâu được áp dụng để phân tích hình ảnh và đưa ra dự đoán về phát hiện đối tượng trong ảnh. Các đối tượng sẽ được định vị trong ảnh bởi một hộp giới hạn (bounding box) với kích thước và tọa độ tâm.



**Hình 1.** Sơ đồ khối của mô hình robot diệt cỏ tự động

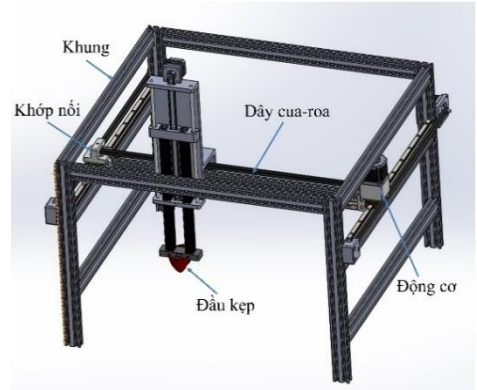
Từ thông tin về tọa độ tâm của đối tượng, tính toán tọa độ của đối tượng trong không gian 3D. Thông tin này được sử dụng để điều khiển các động cơ để dịch chuyển cơ cấu tác vụ đến vị trí cần thiết.

**3. Thiết kế cơ khí**

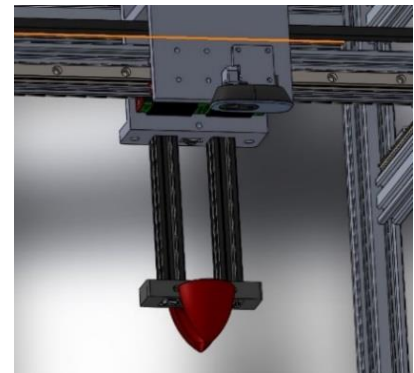
Phần thiết kế cơ khí cho robot bao gồm 3 thành phần chính sau: Khung robot, khối động cơ và truyền động, cơ cấu kẹp. Phần thiết kế và mô phỏng được thực hiện bằng phần mềm SolidWorks (Hình 2).

Khung robot được thiết kế theo hệ Prusa (Hình 2a). Dựa trên các khảo sát thực tế, đưa ra kích thước của robot như sau: 0,75m (D) × 0,7m (R) × 0,7m (C). Phần cơ cấu kẹp có thể dịch chuyển theo 3 trục tự do (Hình 2b). Ba động cơ bước được sử dụng để dịch chuyển cơ cấu đầu kẹp thông qua các dây cua-roa được lắp trên các ray trượt (Hình 2c). Các mạch driver điều khiển động cơ được bố trí dọc lắp trên từng động cơ. Cơ cấu kẹp được điều khiển bằng khí

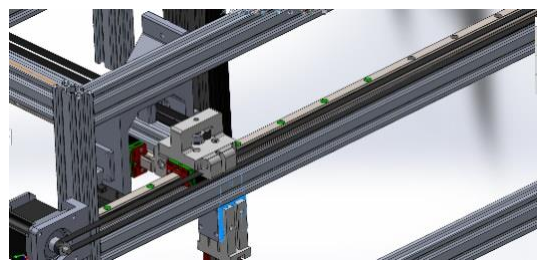
nén, được dự trữ trong bình chứa và được đưa tới cơ cấu chấp hành của đầu kẹp thông qua các van điều khiển. Khi nhận được tín hiệu điều khiển từ vi điều khiển, các van sẽ đóng mở để vận hành cơ cấu kẹp. Chi tiết gắn cơ cấu kẹp được thiết kế để có thể dễ dàng thay thế bởi các cơ cấu tác vụ khác tùy theo mục đích.



a)



b)



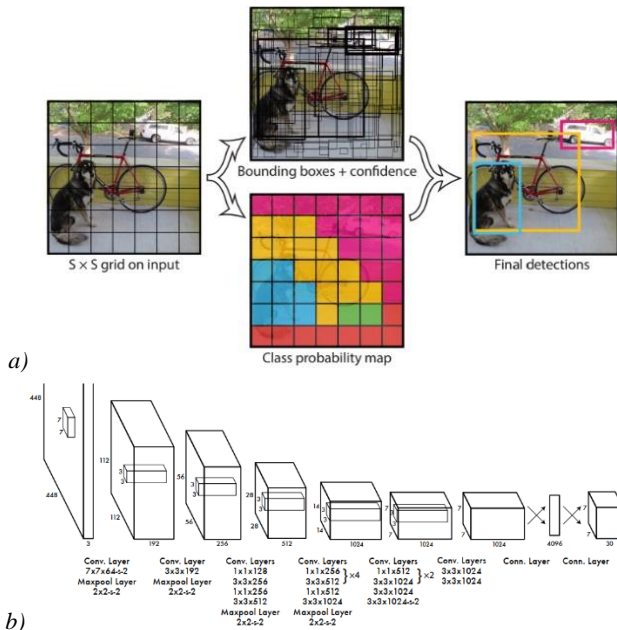
c)

**Hình 2.** Thiết kế 3D: (a) Khung robot; (b) Chi tiết đầu kẹp; (c) Chi tiết ray trượt

**4. Phát hiện và xác định tọa độ 3D của đối tượng**

**4.1. Mô hình phát hiện đối tượng**

Qua khảo sát một số nghiên cứu liên quan [7], nhóm tác giả lựa chọn mô hình YOLO để áp dụng cho khối phát hiện đối tượng. YOLO có ưu điểm về tốc độ xử lý nhanh. Khác với các mô hình dựa trên vùng đề xuất, mô hình YOLO chỉ dựa trên một mạng nơ-ron duy nhất với đầu vào là một ảnh và đầu ra là thông tin về vị trí và kích thước của hộp giới hạn và phân loại của đối tượng phát hiện được. Ảnh đầu vào được chia thành các ô (cell), mỗi ô sẽ dự đoán các thông tin về hộp giới hạn và phân loại của đối tượng (Hình 3a). Mô hình YOLO sử dụng các lớp của mạng nơ-ron tích chập cơ bản để trích xuất các đặc trưng của ảnh. Sau đó, mô hình YOLO thực hiện bài toán hồi quy với dữ liệu vào là các đặc trưng ảnh và dữ liệu ra là các thông tin về hộp giới hạn và phân loại đối tượng thông qua các lớp liên kết đầy đủ (fully-connected layer) (Hình 3b).



**Hình 3.** Mô hình YOLO: (a) Nguyên lý; (b) Kiến trúc mạng [10]

Hàm mất mát của mô hình YOLO bao gồm nhiều thành phần để đánh giá việc dự đoán kích thước và vị trí của hộp giới hạn và việc phân loại đối tượng. Hàm mất mát có dạng như sau [10]:

$$\begin{aligned}
 \mathcal{L} = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
 & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} \left[ \left( \sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 \right. \\
 & \left. + \left( \sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} (C_i - \hat{C}_i)^2 \\
 & + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\
 & + \sum_{i=0}^{S^2} \mathbb{I}_i^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2,
 \end{aligned} \quad (1)$$

với,  $\mathbb{I}_i^{obj}$  thể hiện sự có mặt của đối tượng trong ô thứ  $i$ ;  $\mathbb{I}_{ij}^{obj}$  thể hiện kết quả dự đoán của hộp giới hạn thứ  $j$  trong ô thứ  $i$ ;  $x, y, w, h$  lần lượt là tọa độ tâm và kích thước của hộp giới hạn;  $C$  là độ tin cậy (confidence);  $p(c)$  là xác suất phân loại;  $\lambda_{coord} = 5, \lambda_{noobj} = 0,5$  - trọng số.

#### 4.2. Xác định tọa độ 3D

Do ảnh thu được từ camera chỉ chứa thông tin 2D trong khi không gian thực là 3D nên ta không thể xác định được tọa độ của đối tượng quan tâm trong không gian 3D nếu không có thông tin thêm. Đối với mô hình robot này, ta giữ cố định vị trí của camera so với hệ trục tọa độ gắn với mặt đáy của robot nên dựa trên các thông tin về khoảng cách

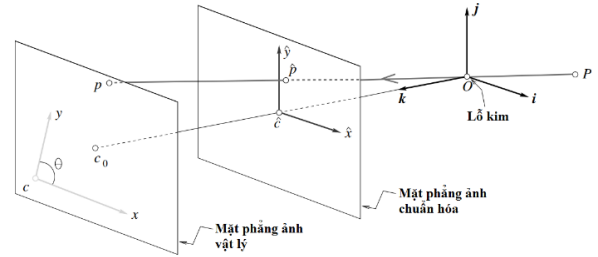
này, ta có thể xác định được tọa độ của các vị trí quan tâm trong không gian 3D từ tọa độ của điểm ảnh. Tuy nhiên, do ảnh hưởng của méo dạng quang học (optical distortions) và sai số trong việc đo khoảng cách nên ta cần có phương pháp hiệu chuẩn (calibration) để tăng độ chính xác của việc xác định tọa độ 3D. Trước hết, chúng ta cần mô hình hóa cho camera và ước lượng các hệ số của mô hình này.

Một camera thông thường có thể mô hình hóa bằng mô hình camera lỗ kim (pinhole camera), tức camera được xem như một hộp kín có một lỗ với khẩu độ nhỏ để ánh sáng từ vật đi qua và tạo ảnh trên cảm biến [13]. Để mô hình hóa chính xác camera thực tế, mô hình camera đầy đủ được bổ sung thêm các méo dạng quang học xuyên tâm (radial distortions) và tiếp tuyến (tangential distortions) sinh ra do các thấu kính của camera.

Để thuận tiện cho việc mô hình hóa camera, ta thêm vào một mặt phẳng ảnh chuẩn hóa (normalized image plane) song song với mặt phẳng ảnh vật lý (physical image plane) và cách lỗ kim một đơn vị khoảng cách (Hình 4). Gọi  $P = (X \ Y \ Z \ 1)^T$ ,  $\hat{p} = (\hat{x} \ \hat{y} \ 1)^T$  và  $p = (x \ y \ 1)^T$  lần lượt là các tọa độ đồng nhất (homogeneous coordinates) của điểm vật trong không gian 3D, điểm ảnh trên mặt phẳng ảnh chuẩn hóa, và điểm ảnh trên mặt phẳng ảnh vật lý. Mô hình camera mô hình hóa phép chiếu từ điểm vật trong không gian 3D đến điểm ảnh trên mặt phẳng ảnh vật lý như sau: Trước tiên, tia sáng từ điểm vật  $P$  tạo ảnh  $\hat{p}$  trên mặt phẳng ảnh chuẩn hóa theo mối quan hệ sau:

$$\hat{p} = \frac{1}{Z} \begin{pmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} P = \frac{1}{Z} \mathcal{T} P, \quad (2)$$

với,  $\mathcal{R}, \mathbf{t}$  - ma trận xoay và vector tịnh tiến để biến đổi điểm  $P$  từ hệ trục tọa độ của không gian 3D sang hệ trục tọa độ của camera;  $\mathcal{T}$  được gọi là ma trận hiệu chuẩn ngoài.



**Hình 4.** Mô hình camera

Do ảnh hưởng của méo dạng quang học xuyên tâm và tiếp tuyến nên tạo độ điểm ảnh bị ảnh hưởng theo biểu thức sau [14]:

$$\hat{p}_d = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \hat{p} + dx, \quad (3)$$

với,  $k_1, k_2, k_3$  là các hệ số méo dạng xuyên tâm, và  $k_4, k_5$  là hệ số méo dạng tiếp tuyến;  $r^2 = \hat{x}^2 + \hat{y}^2$ ;  $dx$  là vector méo dạng tiếp tuyến, có biểu diễn sau:

$$dx = \begin{pmatrix} 2k_4 \hat{x} \hat{y} + k_5 (r^2 + 2\hat{x}^2) \\ k_4 (r^2 + 2\hat{y}^2) + 2k_5 \hat{x} \hat{y} \end{pmatrix}. \quad (4)$$

Cuối cùng, tia sáng tạo ảnh trên mặt phẳng ảnh vật lý. Tọa độ điểm ảnh  $p$  được xác định như sau:

$$p = \begin{pmatrix} \alpha & -\alpha \cot \theta & x_0 \\ 0 & \beta / \sin \theta & y_0 \\ 0 & 0 & 1 \end{pmatrix} \hat{p}_d = \mathcal{K} \hat{p}_d, \quad (5)$$

với  $\alpha, \beta$  - tiêu cự theo phương  $x$  và  $y$  (đơn vị là pixel);  $\theta$  - góc giữa 2 trục của cảm biến (do hệ trục tọa độ của

camera có thể bị méo, tức không hợp góc 90 độ);  $(x_0, y_0)$  – tọa độ của tâm của cảm biến;  $\mathcal{K}$  - ma trận hiệu chuẩn nội.

Để xác định các hệ số của ma trận  $\mathcal{K}$  và  $\mathcal{T}$ , nhóm tác giả thực hiện hiệu chuẩn camera, tương ứng với bài toán tối ưu hóa, với các dữ liệu vào là tọa độ biết trước của các điểm trong không gian 3D và dữ liệu ra là tọa độ các điểm ảnh tương ứng:

$$\operatorname{argmin}_{\Theta} \frac{1}{N} \sum_{i=1}^N \|p_i - \mathcal{F}(P_i, \Theta)\|_2^2, \quad (6)$$

với,  $\Theta$  – là các hệ số của mô hình;  $\{P_i, p_i\}$  là cặp điểm 3D và điểm ảnh;  $\mathcal{F}(\cdot)$  là mô hình camera, được mô tả bởi các phương trình (2-5).

Các thuật toán tối ưu như Levenberg- Marquardt có thể được áp dụng để giải bài toán tối ưu (6). Khi ước lượng được các hệ số của mô hình camera, ta có thể khôi phục được tia sáng tới từ điểm vật dựa trên tọa độ điểm ảnh cho trước bằng cách thực hiện phép chiếu ngược (back-projection) từ phương trình (5) đến (2). Trong đó, để xác định được tọa độ  $\hat{p}$  từ tọa độ  $\hat{p}_d$ , tương ứng với việc khử méo dạng quang học, ta cần giải bài toán tối ưu đơn giản như sau:

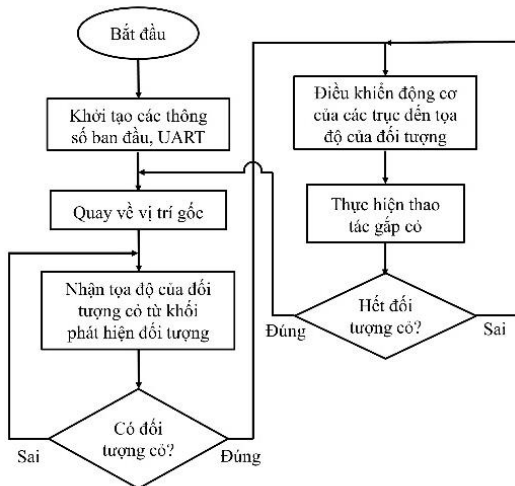
$$\operatorname{argmin}_{\hat{p}} \|\hat{p}_d - \mathcal{S}(\hat{p}, \Omega)\|_2^2, \quad (7)$$

với,  $\mathcal{S}(\cdot)$  – hàm biến đổi được mô tả bởi các phương trình (3) và (4);  $\Omega$  – các hệ số của các phương trình (3) và (4).

Để xác định được tọa độ 3D, nhóm tác giả sử dụng một số điểm dữ liệu 3D với tọa độ biết trước, nằm trên mặt đáy của robot khi thực hiện hiệu chuẩn camera. Nhờ đó, ước lượng được các hệ số của ma trận hiệu chuẩn ngoại  $\mathcal{T}$ , tức thông tin về vị trí tương đối giữa hệ trục tọa độ của mặt đáy (vùng hoạt động của robot) và hệ trục tọa độ của camera.

**5. Điều khiển cơ cấu chấp hành**

Mỗi khung hình được xử lý trên máy tính nhúng Raspberry Pi để phát hiện và xác định tọa độ 3D của các đối tượng cô. Thông tin này được truyền đến khối điều khiển cơ cấu chấp hành. Vi điều khiển ATMega2560 được sử dụng để điều khiển các động cơ bước và các van khí nén. Lưu đồ thuật toán của khối điều khiển cơ cấu chấp hành được mô tả ở Hình 5.

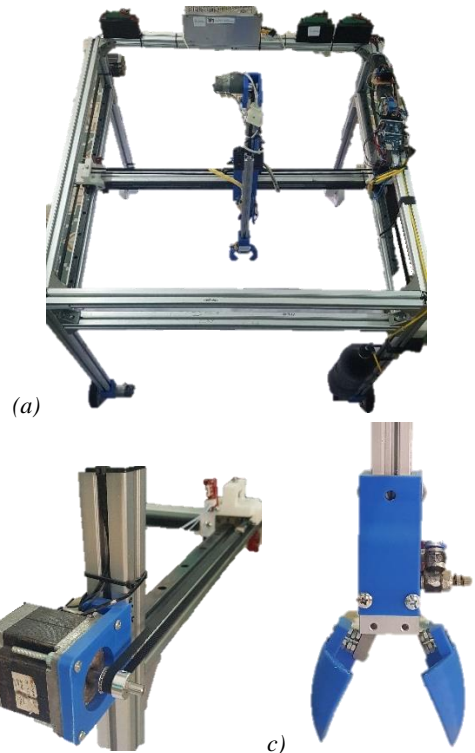


Hình 5. Lưu đồ thuật toán của khối điều khiển cơ cấu chấp hành

**6. Kết quả thực nghiệm**

**6.1. Nguyên mẫu robot**

Từ thiết kế đã được trình bày ở Mục 3, nhóm tác giả đã thi công một nguyên mẫu của robot để thực hiện kiểm tra và đánh giá. Kết quả thi công được thể hiện ở Hình 6. Nhóm tác giả sử dụng các thanh nhôm công nghiệp để làm khung robot. Các chi tiết khớp nối phức tạp và đầu kẹp được chế tạo bằng công nghệ in 3D (Hình 6b&c). Nhóm tác giả sử dụng các động cơ bước với moment xoắn 0.6Nm và các mạch driver TB6600. Camera Logitech C270 được lắp đặt với quang trục vuông góc và cách mặt đáy của robot khoảng 50cm. Với khoảng cách này, hình ảnh thu được từ camera bao quát được vùng làm việc của robot và các đối tượng có kích thước đủ lớn để có thể phát hiện được bằng mô hình YOLO.



Hình 6. (a) Nguyên mẫu robot; (b) Chi tiết động cơ, ray trượt, và khớp nối; (c) Chi tiết cơ cấu kẹp

Kết quả thi công cho thấy, robot có kết cấu vững chắc; động cơ và các cơ cấu truyền động hoạt động ổn; hình ảnh thu được có chất lượng tốt, đảm bảo cho việc phát hiện đối tượng.

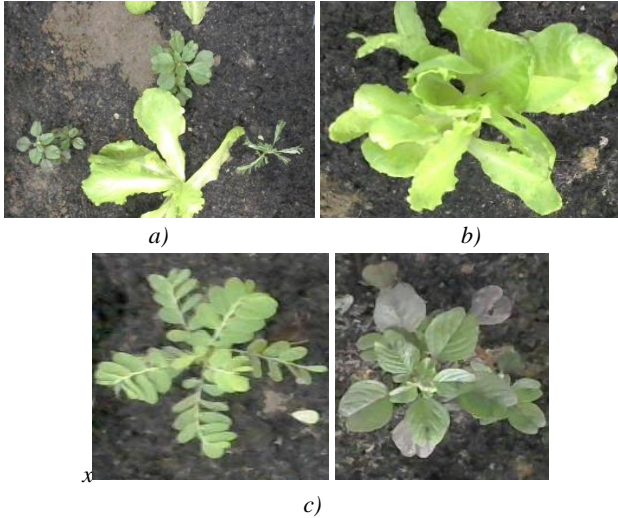
**6.2. Đánh giá mô hình phát hiện đối tượng**

**6.2.1. Cơ sở dữ liệu**

Để huấn luyện mô hình phát hiện đối tượng, nhóm tác giả đã thu thập dữ liệu thực tế từ camera được lắp đặt trên nguyên mẫu robot đã được xây dựng. Với mục đích kiểm tra và đánh giá sơ bộ hiệu quả của mô hình, nhóm tác giả tạo cơ sở dữ liệu nhỏ với một loại cây trồng và hai loại cỏ dại (Hình 7). Đối với mỗi ảnh, tiến hành gán nhãn, bao gồm thông tin về hộp giới hạn và phân loại như đã được trình bày ở Mục 4.1. Cơ sở dữ liệu thu được bao gồm 4531 nhãn. Các kỹ thuật tăng cường dữ liệu (data augmentation) như thay đổi độ sáng, thêm nhiễu, tịnh tiến ảnh, xoay ảnh,... được áp dụng để tăng số lượng mẫu của cơ sở dữ liệu, đồng thời tạo ra sự đa dạng của đối tượng.

6.2.2. Tiêu chí đánh giá và môi trường huấn luyện

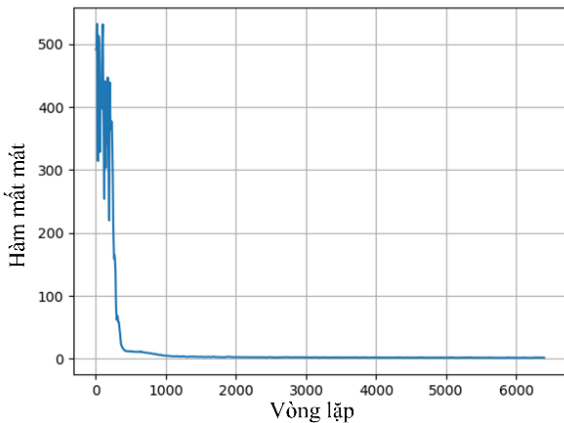
Nhóm tác giả sử dụng các tiêu chí mAP (mean Average Precision) [10] để đánh giá mô hình phát hiện đối tượng. Huấn luyện mô hình trên máy tính có cấu hình như sau: CPU – Intel Core i5-7500 @ 3,4GHz, RAM - 16Gb, GPU - Nvidia GTX GeForce 1070Ti. Ngôn ngữ sử dụng là Python. Áp dụng phương pháp học chuyên tiếp (transfer learning) để huấn luyện mô hình YOLO với cơ sở dữ liệu thu thập được. Tốc độ học được sử dụng là 0,001. Thời gian huấn luyện vào khoảng 8 tiếng.



Hình 7. (a) Hình ảnh thu thập thực tế; (b) Ảnh ví dụ của đối tượng cây trồng; (c) Ảnh ví dụ của đối tượng cỡ đại loại 1 và loại 2

6.2.3. Kết quả huấn luyện và kiểm tra

Sự thay đổi của hàm mất mát theo số vòng lặp được mô tả ở Hình 8. Có thể thấy rằng, hàm mất mát bắt đầu hội tụ từ vòng lặp thứ 1000.



Hình 8. Đồ thị của hàm mất mát theo số vòng lặp

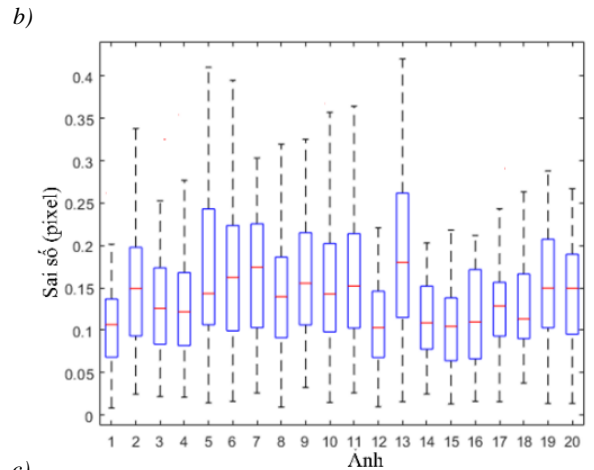
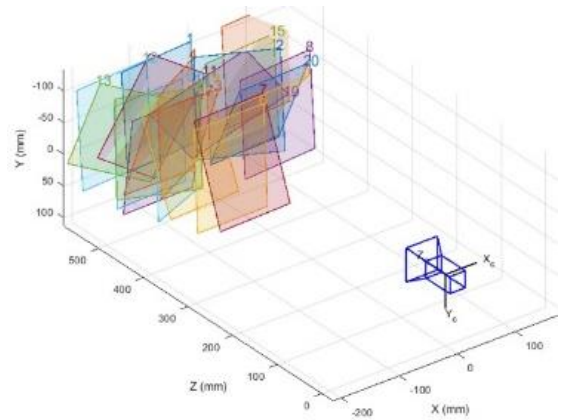
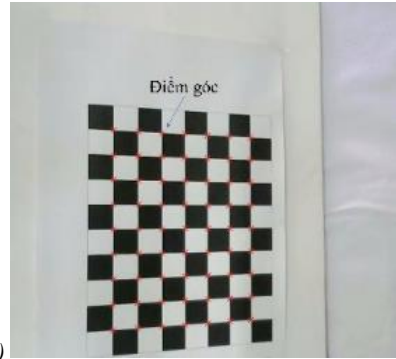
Kết quả mAP được thể hiện ở Bảng 1 cho thấy, độ chính xác của việc phát hiện và phân loại các đối tượng cao. Giá trị mAP trung bình cho tất cả các đối tượng là khoảng 0,95. Kết quả này đảm bảo cho việc tự động phát hiện đối tượng để thực hiện các tác vụ cần thiết.

Bảng 1. Kết quả mAP trên tập kiểm tra

Đối tượng / Tiêu chí	Cây trồng	Cỡ đại 1	Cỡ đại 2
mAP	0,98	0,97	0,91

6.3. Đánh giá sai số xác định tọa độ của đối tượng

6.3.1. Đánh giá sai số hiệu chuẩn camera



Hình 9. (a) Ảnh bảng bàn cờ dùng để hiệu chuẩn camera (các điểm góc được đánh dấu x); (b) Hiện thị trực quan vị trí tương đối giữa các bảng bàn cờ và camera; (c) Sai số khoảng cách giữa điểm ảnh qua phép chiếu và điểm ảnh thật

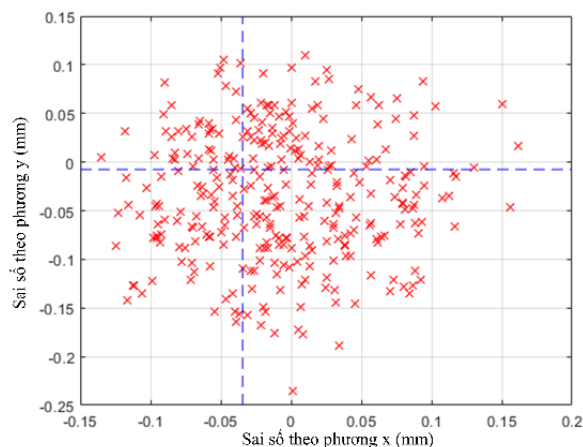
Để hiệu chuẩn camera, nhóm tác giả sử dụng bảng bàn cờ (checkerboard) (Hình 9a), với hệ trục tọa độ cho trước gắn với bảng bàn cờ, chúng ta biết trước tọa độ 3D của các điểm góc (corner) trong bảng. Đồng thời, cũng dễ dàng xác định được tọa độ 2D của các điểm ảnh tương ứng. Với các dữ liệu đầu vào và đầu ra này, thực hiện hiệu chuẩn camera để ước lượng giá trị của các hệ số của camera, và vị trí tương đối giữa camera và bảng bàn cờ.

Hình 9b biểu diễn trực quan vị trí tương đối giữa camera và các bảng bàn cờ. Đặt bảng bàn cờ ở 20 vị trí khác nhau so với camera để kết quả chính xác của việc hiệu chuẩn camera. Sau khi thu được các giá trị ước lượng của

các hệ số của camera và vị trí tương đối, thực hiện phép chiếu từ các điểm góc của từng bảng cờ lên mặt phẳng ảnh dựa trên mô hình của camera (được mô tả bởi các phương trình (2-5)). Từ đó, tính sai số khoảng cách giữa điểm ảnh qua phép chiếu và điểm ảnh thật. Hình 9c mô tả sai số khoảng cách đối với 20 vị trí khác nhau của bảng cờ đối với camera. Có thể thấy rằng, giá trị sai số trung bình đối chỉ vào khoảng 0,14 pixel và giá trị sai số lớn nhất khoảng 0,41 pixel. Kết quả này đảm bảo cho việc xác định tọa độ 3D của vị trí quan tâm được chính xác.

### 6.3.2. Đánh giá sai số xác định tọa độ 3D

Để đánh giá sai số xác định tọa độ 3D, thiết lập các điểm với tọa độ 3D biết trước trên mặt đáy của robot. Vị trí tương đối giữa hệ trục tọa độ của mặt đáy robot và hệ trục tọa độ của camera đã được ước lượng thông qua hiệu chuẩn camera. Từ tọa độ điểm ảnh tương ứng, thực hiện phép chiếu ngược như mô tả ở Mục 4.2 để xác định tọa độ 3D của vị trí quan tâm. Sai số xác định tọa độ của các điểm 3D theo phương  $x$  và phương  $y$  được biểu diễn ở Hình 10. Sai số khoảng cách Euclid trung bình là dưới 0,1 mm. Với sai số nhỏ này, chúng ta có thể thực hiện thao tác kẹp cỏ chính xác.



**Hình 10.** Sai số xác định tọa độ của điểm 3D theo phương  $x$  và phương  $y$ ; Đường nét đứt thể hiện giá trị trung bình theo từng phương

## 7. Kết luận

Trong bài báo này, nhóm tác giả đã trình bày một hệ thống robot diệt cỏ tự động dựa trên phân tích hình ảnh

bằng kỹ thuật học sâu; Đưa ra một thiết kế khung robot theo hệ Prusa với cơ cấu kẹp cỏ nhỏ gọn; Mô hình YOLO được áp dụng để thực hiện phát hiện đối tượng. Nhóm tác giả đã xây dựng một nguyên mẫu của robot theo thiết kế đề xuất. Kết quả thí nghiệm cho thấy, hệ thống có khả năng phát hiện đối tượng và xác định tọa độ 3D của đối tượng với độ chính xác cao. Hướng nghiên cứu tiếp theo sẽ phát triển khối định vị và định hướng để robot có thể tự hành.

## TÀI LIỆU THAM KHẢO

- [1] Sinden J, Jones R, Hester S, et al., *The economic impact of weeds in Australia*. Technical report, CRC for Australian Weed Management, 2004, 1-65.
- [2] Nguyễn Vĩnh Trường, Võ Khánh Ngọc, "Nghiên cứu tính kháng thuốc trừ cỏ chứa hoạt chất pretilachlor của các quần thể cỏ lồng vực ở Quảng Trị", *Tạp chí Khoa học và Công nghệ Nông nghiệp, Trường Đại học Nông Lâm Huế*, 3(1), 2019, 1175-1184.
- [3] Bawden, Owen, et al., "Robot for weed species plant-specific management", *Journal of Field Robotics*, 34(6), 2017, 1179-1199.
- [4] Yu, Jialin, et al., "Weed detection in perennial ryegrass with deep learning convolutional neural network", *Frontiers in Plant Science*, 10, 2019, 1422-1431.
- [5] Chechliński Lukasz, Barbara Siemiątkowska, Michał Majewski, "A System for Weeds and Crops Identification—Reaching over 10 FPS on Raspberry Pi with the Usage of MobileNets, DenseNet and Custom Modifications", *Sensors*, 19(17), 2019, 3787-3801.
- [6] Lottes Philipp, et al., "UAV-based crop and weed classification for smart farming", *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2017.
- [7] Jiao Licheng, et al., "A survey of deep learning-based object detection", *IEEE Access*, 7, 2019, 128837-128868.
- [8] Girshick Ross, et al., "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, 2014.
- [9] Girshick Ross, "Fast R-CNN", *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, 2015, 1440-1448.
- [10] Redmon Joseph, et al., "You only look once: Unified, real-time object detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016, 779-788.
- [11] Farhadi Ali, Joseph Redmon, "Yolov3: An incremental improvement", *Computer Vision and Pattern Recognition*, 2018, 1-5.
- [12] Liu Wei, et al., "SSD: Single shot multibox detector", *European Conference on Computer Vision*, Springer, Cham, 2016, 21-37.
- [13] Forsyth David A., Jean Ponce, *Computer vision: A modern approach*, Pearson, 2012.
- [14] Heikkila Janne, Olli Silvén, "A four-step camera calibration procedure with implicit image correction", *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, IEEE, 1997.