

THUẬT TOÁN DI CHUYỂN THEO ĐỐI TƯỢNG TRÊN MẶT SÀN DỰA TRÊN DÒNG VIDEO NHẬN ĐƯỢC TỪ XE TỰ HÀNH

Nguyễn Hữu Phát*, Nguyễn Tiến Dũng*, Hoàng Bảo Hưng*, Vũ Tiến Đạt†

*Khoa Điện Tử, Trường Điện-Điện Tử, Đại học Bách Khoa Hà Nội

† Khoa Điện Cơ, Trường Đại Học Hải Phòng

Tóm tắt: Mục đích của bài báo này là đề xuất thuật toán xây dựng hệ thống xe tự hành di chuyển theo đối tượng (nhân dân) nhằm giảm bớt công sức và thời gian cho việc tự tay đẩy kéo hoặc mang vác vật. Ý tưởng này có thể ứng dụng cho xe chở hàng trong các nhà kho hoặc cho vali tự động di chuyển theo người dùng. Xe hoạt động theo những bước chính là video được quay liên tục rồi gửi lên server. Từ video đó, server sẽ phát hiện đối tượng có trong khung hình bằng thuật toán YOLO (là đối tượng đã được huấn luyện để phát hiện trước đó) và xác định vị trí tương đối của tâm đối tượng trong khung hình. Cuối cùng nó gửi lệnh xuống xe bám theo đối tượng đó. Do tập dữ liệu huấn luyện còn chưa lớn (392 ảnh) nên độ chính xác của việc phát hiện vật thể còn chưa cao. Tuy nhiên, trong điều kiện thích hợp thì hệ thống có thể xác định chính xác và bám theo đối tượng lên tới 98%. Do đó, nếu tăng tập mẫu ảnh và lựa chọn số lượng bước xử lý thích hợp thì hệ thống hoàn toàn có thể tăng độ chính xác trong những điều kiện ảnh xấu hơn.

Từ khóa: đối tượng; hệ thống xe tự hành; thuật toán; xử lý ảnh; mã hóa.

I. ĐẶT VẤN ĐỀ

Hiện nay, những thuật toán trong lĩnh vực thị giác máy tính đang được phát triển mạnh mẽ và có nhiều ứng dụng như trong giám sát, y tế [1], [10], [11], [12].

Trong lĩnh vực robot di động tự di chuyển trên đường, được trang bị hệ thống máy ảnh stereo, vấn đề quan trọng nhất là chiết tách các mặt phẳng đất và xác định các vật cần để tạo thuận lợi cho phương tiện di chuyển. Trong những năm gần đây, có rất nhiều công việc với các cách tiếp cận tương đối đa dạng để giải quyết câu hỏi. Một số công trình nổi bật [1],[2],[3] sử dụng các thuật toán RANSAC cổ điển để ước tính mặt phẳng đất với độ tin cậy cao. Tuy nhiên, với yêu cầu số lượng lớn các phép toán cần thiết, phương pháp này có thể dẫn đến thời gian chi phí lớn. Hai cách tiếp cận trong [4],[5] áp dụng khái niệm luồng quang và sử dụng dữ liệu đầu vào là chuỗi ảnh màu hoặc ảnh xám đa cấp để có được một số kết quả rất thú vị giới hạn trong môi trường có đặc tính đặc biệt. Trong trường hợp khác, các mặt phẳng đất chứa các mẫu phức tạp, tỷ lệ phát hiện đối tượng chính xác sẽ giảm mạnh. Đối với ứng dụng thời gian thực, các tác giả trong [6] đã trình bày một thuật toán để chỉ cần tìm mặt phẳng đất bằng cách xử lý các dữ liệu bản đồ chênh lệch đầu vào

2D, tuy nhiên những hình ảnh đầu vào được thu thập và được thử nghiệm trong một môi trường trong nhà với một nền tảng đơn giản. Điều này hạn chế khả năng giải quyết các vấn đề trong môi trường thực tế [7],[8].

Xe di chuyển bám theo đối tượng (nhân dân) cũng là một ứng dụng của nó. Trong bài báo [28] đề xuất các biện pháp hiệu suất để đánh giá các thuật toán theo dõi đối tượng bằng cách sử dụng nhân và kích thước đối tượng. Tính hữu ích và hiệu quả của các biện pháp đánh giá đề xuất được thể hiện bằng cách đánh giá hiệu suất của hai thuật toán theo dõi. Trong bài báo [29] đề xuất việc phát hiện và trích xuất cùng một đối tượng chuyển động ở tiền cảnh và nhân phù hợp trong video giám sát nhiều chế độ xem. Mô hình so sánh nền trước nhiều chế độ xem dựa trên phát hiện HOG và phân cụm hệ thống được thiết lập bằng cách sử dụng phương pháp trừ nền, phát hiện tính năng HOG và bình phương sai lệch nhỏ nhất trong phân cụm hệ thống.

Bài báo [30] nghiên cứu cách các đối tượng ghi nhận thông tin được trả về bởi bộ lọc nhiễu Bernoulli được gắn nhãn có thể được sử dụng hiệu quả cho mục đích điều khiển cảm biến. Kết quả mô phỏng cho biết các phương pháp được đề xuất có thể dẫn đến những cải tiến đáng kể về độ chính xác theo dõi của các đối tượng quan tâm, so với việc sử dụng các phương pháp điều khiển cảm biến chung (không chọn lọc). Trong bài báo [31], các tác giả khám phá khái niệm về vòng phản hồi của trình theo dõi-cảm biến có thể thông báo cho quá trình xử lý của cảm biến và điều tra xem liệu nó có thể cải thiện hiệu suất theo dõi nhiều mục tiêu và cho phép phát hiện đối tượng tĩnh trong video hay không.

Trong bài báo [32], các tác giả trình bày một cách tiếp cận tự động truyền các nhãn từ một khung hình duy nhất sang các khung hình tiếp theo. Các tác giả đã chạy thử nghiệm trên chuỗi video điều khiển 101 khung hình tạo ra quy trình cơ sở được gắn nhãn thủ công tương ứng. Kết quả cho thấy hầu hết các pixel được gắn nhãn có thể được truyền chính xác ngay cả sau một trăm khung hình. Tuy nhiên, việc đưa mô hình này vào thực tế còn chưa được rộng rãi vì cho đến nay các thuật toán của thị giác máy tính mới thực sự phát triển mạnh mẽ [11], [12]. Trong điều kiện thực tế, xe di chuyển trên đường phố sẽ gặp phải nhiều vật thể khác nhau, vì vậy ngoài việc các thuật toán cần đáp ứng tốc độ xử lý thời gian thực, còn phải đáp ứng số lượng lớn vật thể có thể nhận diện của hệ thống. Ngoài ra, việc xác định khoảng cách để giữa các vật thể còn chưa được linh hoạt.

Tác giả liên hệ: Nguyễn Hữu Phát,
Email: phat.nguyenhuu@hust.edu.vn

Đến tòa soạn: 10/2021, chỉnh sửa: 01/2022, chấp nhận đăng: 02/2022.

Với bài toán đưa ra, chúng ta không thể sử dụng phương pháp đo truyền thống mà phải sử dụng các thuật toán đo đạc. Thuật toán phổ biến hiện nay là sử dụng hai camera để xác định khoảng cách, tuy nhiên nếu sử dụng hệ thống hai camera, cấu trúc phần cứng trở nên cồng kềnh và khó lắp đặt. Xuất phát từ thực tế đó, trong bài báo này chúng tôi tập trung nghiên cứu vào phần xử lý hình ảnh trong hệ thống xe tự hành từ đó đưa ra một hệ thống mô hình gồm có:

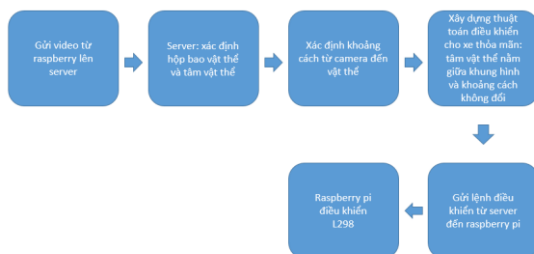
- Xe di chuyển có gắn camera: Xe gồm có Raspberry PI, Module L298 điều khiển động cơ, 2 động cơ, 1 camera; Raspberry pi có nhiệm vụ truyền tải hình ảnh lên server và nhận lệnh từ server điều khiển động cơ di chuyển
- Server: có nhiệm vụ nhận video từ raspberry pi, sau đó chạy thuật toán xử lý và gửi lại lệnh đến raspberry pi.

Phần còn lại của bài báo được trình bày như sau. Trong phần II chúng tôi sẽ khảo sát qua về các hệ thống hiện có. Trong phần III, chúng tôi lần lượt trình bày mô hình và đánh giá kết quả của mô hình đề ra. Cuối cùng, chúng tôi kết luận bài báo trong phần IV.

II. GIẢI PHÁP THỰC HIỆN

A. Cấu trúc hệ thống

Sơ đồ cấu trúc hệ thống được đề xuất như trên hình 1:



Hình 1. Sơ đồ đề xuất quá trình xây dựng hệ thống

Trong bước xác định vật thể, chúng tôi sử dụng thuật toán YOLO [13], [14], [15], [16], [17]. Ưu điểm lớn nhất của thuật toán này so với các thuật toán nhận diện ảnh phổ biến hiện nay như: MobileNet, ResNet,... là tốc độ xử lý nhanh, gần như ngay lập tức nên đáp ứng được yêu cầu đưa ra từ đầu. Với thuật toán này, hình ảnh đầu vào và sẽ được xử lý qua một mạng nơ-ron duy nhất, đầu ra của mạng sẽ là dự đoán tâm và hộp bao vật thể, điều này làm tốc độ xử lý nhanh hơn so với các thuật toán phát hiện vật thể khác, phù hợp với các ứng dụng xử lý ảnh theo thời gian thực.

B. Thu thập tập mẫu và đánh đối tượng

Tập mẫu trong lần thử nghiệm của chúng tôi là các đối tượng cần theo dõi. Trong bài báo này, chúng tôi chỉ đưa ra một vật thể phổ biến trong nhà, với giả định đặt ra là khi xe đến gần đối tượng, xe sẽ giảm tốc độ dần. Yêu cầu đặt ra cho tập mẫu là đủ sự đa dạng, trong nhiều bối cảnh khác nhau, màu sắc, độ sáng ảnh khác nhau để tránh bị overfitting trong quá trình huấn luyện (overfitting: khi huấn luyện kết quả cho ra rất tốt, nhưng khi đưa tập mới thử lại thì kết quả kém)

[12]. Vì vậy chúng tôi đã thu thập tập mẫu là các ảnh thật lấy từ Google và các ảnh tự chụp trong các thời gian, địa điểm, góc chụp khác nhau để làm đa dạng tập mẫu.

Sau khi có tập mẫu sẽ xác định vật thể trong từng ảnh. Bước này phải làm thủ công với sự hỗ trợ của phần mềm labelling. Với mỗi ảnh, ta sẽ chọn tất cả các khu vực có vật thể và sau đó lưu file đối tượng (nhãn dán). Trọng kết quả gồm có: tâm của vật thể, chiều cao, chiều rộng của hộp bao vật thể như trên hình 2.



Hình 2. Ảnh đối tượng đặt trên mặt sàn

C. Huấn luyện tập mô hình

Sau quá trình huấn luyện, kết quả sẽ thu được một bộ trọng số và một mô hình mạng (mô hình gồm các thông số về số lượng nơ-ron, giá trị tỉ lệ học: learning rate, số lớp,...) các kết quả sau sẽ được xây dựng dựa theo những bộ trọng số và mô hình trước đó đã được thực hiện và đã được công bố miễn phí trên Darknet [18], [19].

Để xây dựng mô hình này, chúng tôi sử dụng bộ công cụ darkflow, gồm các thư viện hỗ trợ huấn luyện mạng YOLO. YOLO là một trong những thuật toán object detection mạnh nhất hiện nay. Mục tiêu của mô hình không chỉ là dự báo nhãn cho vật thể như các bài toán classification mà nó còn xác định location của vật thể. Do đó YOLO có thể phát hiện được nhiều vật thể có nhãn khác nhau trong một bức ảnh thay vì chỉ phân loại duy nhất một nhãn cho một bức ảnh.

YOLO nhìn toàn bộ hình ảnh cùng một lúc; do đó, sự phát hiện của nó có được bởi toàn bộ thông tin trong hình ảnh. YOLO sử dụng mạng đơn, không giống như các máy dò mạng nơ-ron phức tạp khu vực (R-CNN và Fast-RCNN) cần hàng nghìn khu vực cho một hình

ảnh. Mô hình có được các tính năng đầu vào hình ảnh và chia nó thành lưới $S \times S$. Sau đó, nó trích xuất dữ liệu từ mỗi lưới trong hình ảnh đầu vào và sử dụng điểm tin cậy của đối tượng được phát hiện để tạo ra hộp hình chữ nhật. Mỗi ô dự đoán hộp giới hạn và điểm tin cậy. Hộp giới hạn chứa năm tham số dự đoán, được xác định bởi (x, y, w, h) và giá trị điểm tin cậy, trong đó (x, y) tọa độ đại diện cho tâm của hộp giới hạn và (h, w) phản ánh chiều cao và chiều rộng của toàn bộ hình ảnh. Điểm tin cậy đại diện cho phép đo mức độ tự tin của kiến trúc mà hộp chứa đối tượng cần dự đoán. Kết quả ví dụ thử nghiệm như trên hình 3.

```
Dataset size: 392
Dataset of 392 instance(s)
Training statistics:
  Learning rate : 1e-05
  Batch size    : 8
  Epoch number  : 200
  Backup every  : 2000
step 1 - loss 106.33396911621094 - moving ave loss 106.33396911621094
step 2 - loss 107.127967183447266 - moving ave loss 106.41336898803712
```

Hình 3. Quá trình huấn luyện.

III. KẾT QUẢ ĐẠT ĐƯỢC

A. Tính thông số liên quan đến hộp bao vật thể

Sau khi quá trình huấn luyện kết thúc ta có bộ trọng số và cấu hình mạng mới. Kết quả như trên hình 4 và hình 5.

[illegible]

```
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 30.1236 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.7591 cm
Distance from cameras to object : 29.9403 cm
Distance from cameras to object : 29.9403 cm
```

Hình 4. Kết quả phát hiện vật thể khoảng cách a) 20 cm và b) 30 cm

```
{ 'label': 'hung', 'confidence': 0.58272636, 'bottomright': { 'x': 349, 'y': 379}, 'topleft': { 'x': 289, 'y': 174}}
{ 'label': 'hung', 'confidence': 0.58272636, 'bottomright': { 'x': 349, 'y': 379}, 'topleft': { 'x': 289, 'y': 174}}
{ 'label': 'hung', 'confidence': 0.58272636, 'bottomright': { 'x': 349, 'y': 379}, 'topleft': { 'x': 289, 'y': 174}}
{ 'label': 'hung', 'confidence': 0.517604, 'bottomright': { 'x': 355, 'y': 355}, 'topleft': { 'x': 283, 'y': 124}}
{ 'label': 'hung', 'confidence': 0.517604, 'bottomright': { 'x': 355, 'y': 355}, 'topleft': { 'x': 283, 'y': 124}}
{ 'label': 'hung', 'confidence': 0.80468816, 'bottomright': { 'x': 352, 'y': 374}, 'topleft': { 'x': 287, 'y': 178}}
{ 'label': 'hung', 'confidence': 0.83060914, 'bottomright': { 'x': 351, 'y': 372}, 'topleft': { 'x': 287, 'y': 180}}
{ 'label': 'hung', 'confidence': 0.75533813, 'bottomright': { 'x': 352, 'y': 377}, 'topleft': { 'x': 286, 'y': 175}}
{ 'label': 'hung', 'confidence': 0.75533813, 'bottomright': { 'x': 352, 'y': 377}, 'topleft': { 'x': 286, 'y': 175}}
{ 'label': 'hung', 'confidence': 0.87302387, 'bottomright': { 'x': 298, 'y': 350}, 'topleft': { 'x': 242, 'y': 202}}
{ 'label': 'hung', 'confidence': 0.87302387, 'bottomright': { 'x': 298, 'y': 350}, 'topleft': { 'x': 242, 'y': 202}}
{ 'label': 'hung', 'confidence': 0.6170559, 'bottomright': { 'x': 253, 'y': 385}, 'topleft': { 'x': 188, 'y': 167}}
{ 'label': 'hung', 'confidence': 0.6422932, 'bottomright': { 'x': 252, 'y': 382}, 'topleft': { 'x': 189, 'y': 170}}
{ 'label': 'hung', 'confidence': 0.6422932, 'bottomright': { 'x': 252, 'y': 382}, 'topleft': { 'x': 189, 'y': 170}}
{ 'label': 'hung', 'confidence': 0.6219283, 'bottomright': { 'x': 257, 'y': 374}, 'topleft': { 'x': 185, 'y': 178}}
{ 'label': 'hung', 'confidence': 0.70332456, 'bottomright': { 'x': 255, 'y': 381}, 'topleft': { 'x': 187, 'y': 171}}
```

Hình 5. Thông số liên quan đến hộp bao vật thể.

Trong đó: ‘label’ là tên vật thể, ‘confidence’ là độ chính xác của dự đoán, ‘bottomright’ là tọa độ góc dưới bên phải của hộp, ‘topleft’ là tọa độ góc trên bên trái của hộp (gốc tọa độ nằm ở góc trên bên trái của khung hình).

Từ các tọa độ trên ta xác định được tâm và kích thước của hộp. Ví dụ ‘bottomright’: (255,381), ‘topleft’: (187, 171) ta có tâm C của hộp là:

$$C\left(\frac{255+187}{2}; \frac{381+171}{2}\right) = C(221; 276) \quad (1)$$

B. Tính khoảng cách từ camera đến vật thể

Có 2 phương án để tính khoảng cách từ camera đến vật thể: i) sử dụng 2 camera và đo đặc dựa vào các tính toán vật lý (dựa vào tiêu cự và kích thước ảnh trên khung hình); ii) sử dụng một camera để đo khoảng cách (kết quả tương đối). Trong bài báo này đề xuất phương án sử dụng 1 camera để giảm độ phức tạp cho việc lập trình và vì kết quả dự đoán vật thể từ 2 camera không khớp nhau 100%.

Các bước thực hiện như sau:

Bước 1: Xác định được hộp bao vật thể

Bước 2: Tính tỉ lệ (T) chiều rộng của hộp bao/ chiều rộng của khung hình

Bước 3: Tính khoảng cách S bằng cách nhân T với hệ số cố định a (a được xác định dựa vào đo đặc khoảng cách của vật tới camera trong thực tế và tỉ số T tương ứng khi đó (hàm biểu diễn là hàm tuyến tính). Kết quả thể hiện trên hình 6.

Ưu điểm: dễ thực hiện, giảm thời gian xử lý và lập trình

Nhược điểm: yêu cầu đối tượng (nhân dân) phải có kích thước cố định và kết quả chỉ tương đối do góc quay camera ảnh hưởng đến dự đoán (kết quả tốt nhất khi nhìn thẳng). Tuy nhiên, với yêu cầu của xe thì việc sai số như vậy vẫn chấp nhận được.

C. Xây dựng thuật toán điều khiển xe

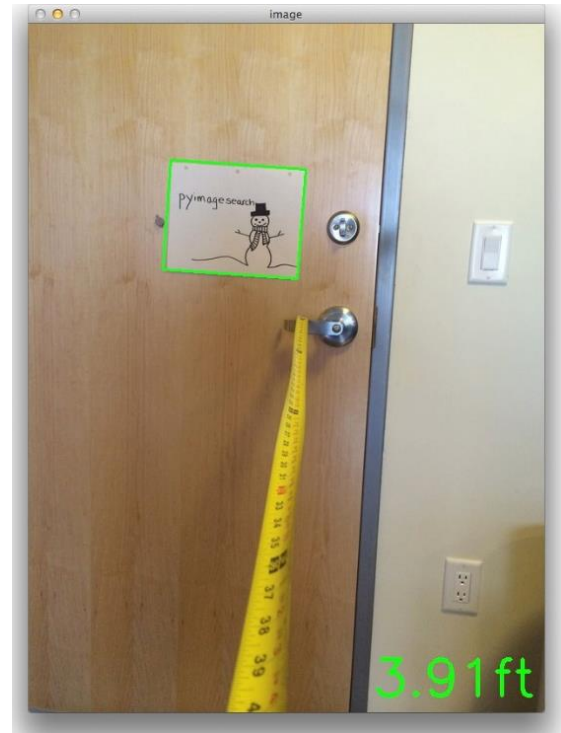
Bước 1: Vị trí cân bằng của xe là khi tọa độ tâm của vật thể có hoành độ nằm giữa khung hình.

Bước 2: Khi đối tượng (nhân dân) di chuyển sang trái, hoành độ của vật thể nhỏ dần. Khi đó ta sẽ điều khiển bánh xe (bánh trái có tốc độ 0, bánh phải có tốc độ 0+x) để xe quay sang trái sao để hoành độ của tâm sẽ dịch về vị trí cân bằng (camera gắn cố định trên xe).

Bước 3: Khi đối tượng (nhân dân) có khoảng cách lớn hơn 1 ngưỡng đặt trước, xe sẽ di chuyển lên phía trước (tránh trái có tốc độ b+0, bánh phải có tốc độ b+0+x).

Bước 4: Khi đối tượng (nhân dân) có khoảng cách nhỏ hơn ngưỡng đặt trước xe dừng lại.

Bước 5: Gửi lệnh từ server đến raspberry pi và điều khiển L298 bằng raspberry pi sử dụng các chân GPIO của Pi để điều khiển.



Hình 6. Tính toán khoảng cách.

D. Thảo luận

Một trong những phương pháp truyền thống có thể đo được khoảng cách xa và tương đối chính xác là phương pháp sử dụng tia laze. So với phương pháp được đưa ra trong đề tài, chúng tôi đưa ra một số nhận xét như bảng 1:

Bảng 1. So sánh phương pháp đề xuất với các phương pháp khác

Thông số	Phương pháp truyền thống dùng Laze [20] ÷ [27]	Phương pháp đề xuất sử dụng mạng YOLO
Độ chính xác	lên đến 99%	70 - 98%
Độ trễ	< 0.5s	< 0.5s
Khoảng cách đo	< 30m	< 80m
Chi phí	dưới 2 triệu đồng	dưới 2 triệu đồng

Nhận xét: Độ chính xác của hệ thống đo qua thuật toán có thể được tăng lên với lượng đầu vào cho phần học máy nhiều và chất lượng hơn. Bên cạnh đó phương

pháp đề xuất cũng đo được với khoảng cách xa hơn với độ trễ như nhau.

IV. KẾT LUẬN

Bài báo nghiên cứu việc xây dựng hệ thống xe di chuyển theo đối tượng (nhân dân). Ở đây chúng tôi mô phỏng quá trình tập hợp dữ liệu huấn luyện và đánh giá kết quả. Hướng tiếp theo chúng tôi sẽ thực hiện:

- Mô phỏng quá trình tập hợp dữ liệu cho toàn bộ xe cùng một thời điểm.
- Đánh giá mức tiêu thụ năng lượng của thuật toán và hệ thống.
- Xử lý song song nhiều đối tượng (nhân dân).

LỜI CẢM ƠN

Nghiên cứu này được thực hiện trong khuôn khổ đề tài do Trường Đại Học Bách Khoa, Việt Nam tài trợ với tiêu đề **Nghiên cứu phát triển thuật toán phát hiện mặt sàn dựa trên dòng video độ sâu từ hệ thống cảm biến Realsense ứng dụng hỗ trợ dẫn đường cho xe tự hành trong nhà** theo đề tài cấp trường mã số **T2021-TT-004**. Cảm ơn Trường ĐHBK đã tài trợ trong quá trình thực hiện bài báo này.

TÀI LIỆU THAM KHẢO

- [1] Z. Su, C. Wang, Y. Li and X. Wu, "A method of cliff detection in robot navigation based on multi-sensor," 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2020, pp. 152-157, doi: 10.1109/RCAR49640.2020.9303038.
- [2] Xiao Hu, Rodriguez, Geppeth: A multi-modal system for road detection and segmentation. Intelligent Vehicles Symposium Proceedings, IEEE, pp.1365 – 1370, (2014).
- [3] Anders Hast, Johan Nysjö, Andrea Marchetti: Optimal RANSAC – Towards a Repeatable Algorithm for Finding the Optimal Set. Journal of WSCG, pp. 21–30, (2013).
- [4] R. Verma and A. K. Verma, "An Efficient Clustering Algorithm to Simultaneously Detect Multiple Planes in a Point Cloud," 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), 2020, pp. 154-159, doi: 10.1109/ICETCE48199.2020.9091735.
- [5] Arshad Jamal, Praveen Mishra, Subrata Rakshit, Abhishekk Kumar Singh, Manish Kumar: Real-time ground plane segmentation and obstacle detection for mobile robot navigation. Emerging Trends in Robotics and Communication Technologies (INTERACT), pp. 314 – 317, (2010).
- [6] Buis, Jan Maarten, and Lode De Herdt. "Obstacle and Cliff Detection for Robotics Applications Using Minutized Sonar and IR Distance Triangulation: Environment Observation Module For The Deci Zebro." (2017).
- [7] P. K. Rath, A. Ramirez-Serrano and D. K. Pratihari, "Moving object detection for humanoid navigation in cluttered dynamic indoor environments using a confidence tracking approach," 2020 Digital Image Computing: Techniques and Applications (DICTA), 2020, pp. 1-7, doi: 10.1109/DICTA51227.2020.9363413.
- [8] CheeWay Teoh, ChingSeong Tan, Yong Chai Tan: Ground plane detection for autonomous vehicle in rainforest terrain. Sustainable Utilization and Development in Engineering and Technology (STUDENT), IEEE Conference on, pp. 7 – 12 (2010).
- [9] R. Mueid, L. Christopher and R. Tian, "Vehicle-pedestrian dynamic interaction through tractography of relative movements and articulated pedestrian pose estimation," IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, 2016, pp. 1-6, 2016.
- [10] E. Yurtsever, J. Lambert, A. Carballo and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," in IEEE Access, vol. 8, pp. 58443-58469, 2020.
- [11] S. J. Henderson and S. Feiner, "Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret," 8th IEEE International Symposium on Mixed and Augmented Reality, Orlando, FL, 2009, pp. 135-144, 2009.
- [12] N. O. Alsrehin, A. F. Klaib and A. Magableh, "Intelligent Transportation and Control Systems Using Data Mining and Machine Learning Techniques: A Comprehensive Study," In IEEE Access, vol. 7, pp. 49830-49857, 2019.
- [13] Du, Juan, "Understanding of Object Detection Based on CNN Family and YOLO," Journal of Physics: Conference Series, 2018.
- [14] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, 2016.
- [15] R. Huang, J. Pedoeem and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 2503-2510, 2018.
- [16] A. Ćorović, V. Ilić, S. Đurić, M. Marijan and B. Pavković, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," 26th Telecommunications Forum (TELFOR), Belgrade, pp. 1-4, 2018.
- [17] S. Wang, L. Niu and N. Li, "Research on Image Recognition of Insulators Based on YOLO Algorithm," International Conference on Power System Technology (POWERCON), Guangzhou, 2018, pp. 3871-3874, 2018.
- [18] J. Won, D. Lee, K. Lee and C. Lin, "An Improved YOLOv3-based Neural Network for De-identification Technology," 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Jeju, Korea (South), pp. 1-4, 2019.
- [19] S. Mahurkar, "Integrating YOLO Object Detection with Augmented Reality for iOS Apps," 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York City, NY, USA, pp. 585-589, 2018.
- [20] Suh, Young, "Laser Sensors for Displacement, Distance and Position," Sensors. 19. 1924. 10.3390/s19081924, 2019.
- [21] C. C. Chang, et al., "The Study of Ultrasonic Distance Measurement Device for a Teleoperated Robotic Manipulator System," in OCEANS, MTS/IEEE Conference and Exhibition, vol. 2, pp. 1054–1057, 2001.
- [22] H. Kim, et al., "Distance Measurement Using a Single Camera with a Rotating Mirror," International Journal of Control, Automation, and Systems, vol/issue: 3(4), pp. 542-551, 2005.
- [23] Kh. A. Rahman, et al., "Person to Camera Distance Measurement Based on Eye-Distance," in IEEE 3rd International Conference on Multimedia and Ubiquitous Engineering, MUE '09, pp. 137–141, 2009.
- [24] T. Mohammad, "Using Ultrasonic and Infrared Sensors for Distance Measurement," International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering, vol/issue: 3(3), pp. 267-272, 2009.
- [25] L. Xiaoming, et al., "Real-Time Distance Measurement Using a Modified Camera," in IEEE Sensors Applications Symposium (SAS), 23-25 Feb. 2010, Limerick, pp. 54–58, 2010.
- [26] A. K. Shrivastava, et al., "Distance Measurement of an Object or Obstacle by Ultrasound Sensors using P89C51RD2," International Journal of Computer Theory and Engineering, vol/issue: 2(1), pp. 1793-8201, 2010.

- [27] W. Jinjin, et al., "Range Resolution of Ultrasonic Distance Measurement Using Single Bit Cross Correlation for Robots," in IEEE International Conference on Information and Automation (ICIA), 20-23 June 2010, Harbin, China, pp. 917-923, 2010.
- [28] J. Popoola and A. Amer, "Performance evaluation for tracking algorithms using object labels," 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 2008, pp. 733-736, doi: 10.1109/ICASSP.2008.4517714.
- [29] W. Zhou, Z. Li and P. Gao, "Research on Moving Object Detection and Matching Technology in Multi-Angle Monitoring Video," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 741-744, doi: 10.1109/ITAIC.2019.8785803.
- [30] S. Panicker, A. K. Gostar, A. Bab-Haidashar and R. Hoseinnezhad, "Sensor Control for Selective Object Tracking Using Labeled Multi-Bernoulli Filter," 2018 21st International Conference on Information Fusion (FUSION), Cambridge, 2018, pp. 2218-2224, doi: 10.23919/ICIF.2018.8455829.
- [31] K. Ingersoll, P. C. Niedfeldt and R. W. Beard, "Multiple target tracking and stationary object detection in video with Recursive-RANSAC and tracker-sensor feedback," 2015 International Conference on Unmanned Aircraft Systems (ICUAS), Denver, CO, 2015, pp. 1320-1329, doi: 10.1109/ICUAS.2015.7152426.
- [32] J. Fauqueur, G. Brostow and R. Cipolla, "Assisted Video Object Labeling By Joint Tracking of Regions and Keypoints," 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, 2007, pp. 1-7, doi: 10.1109/ICCV.2007.4409124.

PROPOSED OBJECTIVES ON THE FLOOR ORGANIZATION BASED ON VIDEO RECEIVED FROM AUTOMATIC VEHICLES

Abstract: The goal of this paper is to propose algorithm to build a car system for moving by object (label); moving vehicles following the sticker is an idea to reduce effort and time for pushing or pulling and carrying the carts by hand. This idea can be applied to cargo trucks in warehouses or for suitcases that automatically move with the user. The vehicle works in the main steps: Video is continuously recorded and then sent to the server, from that video, the server will detect the sticker in the frame by YOLO algorithm (the sticker has been trained to detect) and determine the relative position of the center of the label in the frame, finally sending the order down to the vehicle following that sticker. Because the training data set is small (392 images), the detection accuracy of the object is not high, but under the right conditions it can be determined up to 96%. By increasing the sample set and choosing the appropriate number of steps, it is possible to increase accuracy in worse image conditions.

Keywords: object; label; automatic vehicle; YOLO algorithm; label processing; encoding.



Nguyen Huu Phat, nhận bằng kỹ sư (2003), thạc sĩ (2005) ngành Điện tử và Viễn thông tại Đại học Bách Khoa Hà Nội (HUST), Việt Nam và bằng tiến sĩ (2012) về Khoa học Máy tính tại Viện Công nghệ Shibaura, Nhật Bản. Hiện tại, đang là giảng viên tại Viện Điện tử Viễn thông, HUST, Việt Nam. Năm 2022, ông được Hội đồng chức danh giáo sư Nhà nước công nhận đạt chuẩn PGS và được Trường Đại học Bách Khoa bổ nhiệm học hàm PGS. Các nghiên cứu gồm xử lý hình ảnh và video, mạng không dây, big data, hệ thống giao thông thông minh (ITS), và internet của vạn vật (IoT).



Nguyen Tien Dung, nhận bằng đại học năm 1996 chuyên ngành Công nghệ Điện tử tại Trường Đại học kỹ thuật Sô-phi-a, Bulgaria; nhận bằng thạc sĩ năm 2000 chuyên ngành điện tử viễn thông tại Trường Đại học Bách Khoa Hà Nội, Việt Nam; nhận bằng tiến sĩ năm 2005 chuyên ngành Điện tử tại Trường Đại học Quốc gia Kyungpook, Hàn quốc. Từ năm 1997, ông bắt đầu làm việc tại Khoa Điện tử Viễn thông, nay là Viện Điện tử Viễn thông, trường Đại học Bách Khoa Hà Nội. Năm 2010, ông được Hội đồng chức danh giáo sư Nhà nước công nhận đạt chuẩn PGS và được Trường Đại học Bách Khoa bổ nhiệm học hàm PGS. Hướng nghiên cứu chính của PGS. TS Nguyễn Tiến Dũng tập trung vào xử lý ảnh và video, ứng dụng trong các lĩnh vực điện tử y sinh, giao thông thông minh, hệ thống định vị dẫn đường hỗ trợ robot cứu hộ cứu nạn ...



Hoang Bao Hung, là sinh viên Viện Điện tử Viễn thông, Trường Đại Học Bách Khoa Hà Nội. Hướng nghiên cứu gồm xử lý hình ảnh và video kỹ thuật số và các ứng dụng nhà thông minh.



Vu Tien Dat, là giảng viên khoa Điện Cơ, Trường Đại Học Hải Phòng. Hướng nghiên cứu gồm đo lường, điều khiển tự động, tin học công nghiệp và xử lý tín hiệu.