ƯỚC LƯỢNG KHOẢNG CÁCH VẬT THỂ TỪ ẢNH CAMERA ĐƠN TRONG ĐIỀU KIỆN GÓC PITCH THAY ĐỔI BẰNG HỌC SÂU VÀ MÔ HÌNH CAMERA LỖ KIM

Trần Khắc Trí, Lâm Thị Hiền, Bùi Liên Tuệ Nhi, Nguyễn Minh Hải*

Khoa Vật lý, Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam

5001105048@student.hcmue.edu.vn, 5001105012@student.hcmue.edu.vn, 5001105030@student.hcmue.edu.vn, *hainm@hcmue.edu.vn*

TÓM TẮT— Nghiên cứu này giải quyết bài toán ước lượng khoảng cách tới phương tiện phía trước trong các hệ thống hỗ trợ lái xe nâng cao (ADAS) sử dụng camera đơn. Thách thức cốt lõi của các mô hình hình học truyền thống nằm ở sai số tích lũy do sự thay đổi của góc pitch camera trong điều kiện vận hành thực tế. Chúng tôi đề xuất một quy trình tích hợp kết hợp mạng học sâu YOLOv8s để nhận diện đối tượng và mô hình camera lỗ kim để tính toán khoảng cách từ ảnh đơn đầu vào. Trong nghiên cứu này, chúng tôi trích xuất các đặc trưng hình học từ mặt đường thông qua kỹ thuật LSD (Line Segment Detector) và thuật toán RANSAC bền vững để xác định điểm tụ (Vanishing Point), từ đó cho phép bù sai số góc pitch mà không cần các cảm biến IMU đắt tiền. Kết quả thực nghiệm khẳng định tính hiệu quả của mô hình với sai số trung bình đạt 10,53% trong phạm vi 45,00 m. Giải pháp này không chỉ đảm bảo độ tin cậy về mặt đo lường mà còn tối ưu hóa hiệu suất tính toán, mở ra hướng đi mới cho các ứng dụng thị giác máy tính chi phí thấp trên hệ thống cảnh báo va chạm giao thông và phương tiện tự hành.

Từ khóa— hiệu chỉnh camera, phát hiện đối tượng, góc pitch, ước lượng khoảng cách, thị giác máy tính.

I. GIỚI THIỆU

Sự gia tăng các vụ tai nạn giao thông trên toàn cầu đã khiến an toàn đường bộ trở thành vấn đề y tế công cộng cấp bách. Theo Tổ chức Y tế Thế giới (WHO), mỗi năm có khoảng 1,45 triệu người tử vong và hơn 50 triệu người bị thương do tai nạn giao thông. Dự đoán đến năm 2030, đây sẽ trở thành nguyên nhân tử vong đứng thứ năm toàn cầu [1]. Đặc biệt, va chạm từ phía sau chiếm tỷ trọng lớn nhất, đạt 32,3% tổng số vụ tai nạn tại Hoa Kỳ (2018) và 36,5% tại Trung Quốc (2017).

Trước bối cảnh đó, việc xác định chính xác khoảng cách tới phương tiện phía trước là yếu tố then chốt để kích hoạt các hệ thống hỗ trợ lái xe nâng cao (ADAS) nhằm đưa ra cảnh báo kịp thời [2]. Tuy nhiên, một thách thức kỹ thuật lớn hiện nay là sự biến động của góc pitch camera trong thực tế. Các mô hình hiện tại thường giả định một góc lắp đặt cố định, dẫn đến sai số nghiêm trọng khi phương tiện di chuyển trong điều kiện thực tế. Sự thay đổi dù nhỏ của góc pitch cũng gây ra sự sai lệch đáng kể đến kết quả của việc ước lượng khoảng cách.

Nghiên cứu này đề xuất một khung nghiên cứu toàn diện cho bài toán ước lượng khoảng cách từ ảnh đơn. Những đóng góp chính của nghiên cứu gồm:

- Sử dụng phương pháp của Zhang để xác định chính xác các thông số nội tại của camera tạo tiền đề cho việc ước lượng khoảng cách.
- Phân tích hiệu năng các mô hình nhận diện đối tượng và sử dụng YOLOv8s để xác định điểm tiếp xúc với mặt đường.
- Xác định góc pitch dựa trên cấu trúc tuyến tính của mặt đường bằng kỹ thuật LSD và RANSAC, giúp hệ thống vận hành bền vững trong nhiều điều kiện bối cảnh khác nhau.
- Xây dựng thuật toán ước lượng khoảng cách dựa trên mô hình camera lỗ kim trong điều kiện góc pitch thay đổi.
- Đánh giá thực nghiệm về độ chính xác và hiệu quả tính toán, chứng minh khả năng ứng dụng thực tiễn của giải pháp chi phí thấp.

Trong phần II, các nghiên cứu liên quan đến vấn đề ước lượng khoảng cách sẽ được phân tích. Phần III nói về phương pháp đề xuất bao gồm các giai đoạn cần xử lý: sơ đồ tổng quan, thu thập dữ liệu, hiệu chuẩn camera, nhận diện đối tượng, xác định góc pitch và ước lượng khoảng cách. Tiếp đến, phần IV sẽ tiến hành và phân tích kết quả thực nghiệm. Cuối cùng, phần V sẽ nhận xét tổng quan và đưa ra định hướng nghiên cứu trong tương lai.

II. NGHIÊN CỨU LIÊN QUAN

Ước lượng khoảng cách là thành phần thiết yếu trong các hoạt động tự hành và robot, hỗ trợ đắc lực cho việc hiểu cảnh vật xung quanh và tránh chướng ngại vật [3]. Các phương pháp phát triển cho nhiệm vụ này có thể được chia thành ba nhóm chính: phương pháp chủ động, phương pháp dựa trên hình học và phương pháp dựa trên mạng học sâu.

Nhóm các phương pháp chủ động dựa vào camera RGB-D, LIDAR, Radar hoặc thiết bị siêu âm để trực tiếp lấy thông tin vị trí của các vật thể trong cảnh [4]. Radar được sử dụng rộng rãi nhờ khả năng hoạt động ổn định trong mọi điều kiện ánh sáng và thời tiết, tuy nhiên, nó cung cấp dữ liệu độ phân giải thấp và khó mô tả chính xác hình

*Corresponding author: Nguyễn Minh Hải (*hainm@hcmue.edu.vn*)

dạng vật thể [5]. Ngược lại, LiDAR cho phép xác định khoảng cách với sai số cực thấp và tái tạo mô hình 3D chi tiết thông qua hàng ngàn điểm laser [6]. Mặc dù có độ tin cậy cao, nhược điểm lớn nhất của các cảm biến này là chi phí sản xuất đắt đỏ và yêu cầu tài nguyên tính toán lớn, gây khó khăn cho việc triển khai trên các phương tiện phổ thông hoặc thiết bị nhúng hạn chế.

Nhóm phương pháp dựa trên hình học khôi phục cấu trúc không gian của cảnh từ hình ảnh 2D thông qua các ràng buộc hình học. Kỹ thuật Structure from Motion (SfM) và hệ thống camera stereo [7], [8] đã được nghiên cứu rộng rãi nhưng thường gặp vấn đề về độ phức tạp tính toán hoặc sự bất định về tỷ lệ khi chỉ sử dụng một camera đơn. Nghiên cứu của Stein và cộng sự [9] đã chứng minh khả năng ước lượng khoảng cách từ camera đơn cho hệ thống ACC dựa trên quy luật phối cảnh, với điều kiện camera đặt ở độ cao cố định. Tiếp nối hướng đi này, Park và cộng sự [10] đề xuất ước lượng đường chân trời ảo từ vị trí phương tiện để tính toán khoảng cách ngay cả khi độ dốc mặt đường thay đổi. Đáng chú ý, Chen và cộng sự [11] đã sử dụng các ràng buộc làn đường để xác định tư thế camera, giúp giảm bớt việc hiệu chuẩn thủ công. Tuy nhiên, phương pháp này phụ thuộc mật thiết vào các dấu hiệu hạ tầng đường bộ và chưa giải quyết triệt để bài toán bù sai số động do sự biến đổi góc pitch tức thời của phương tiện. Khoảng trống nghiên cứu về một phương pháp tích hợp hoàn chỉnh, từ xử lý dữ liệu đầu vào cho đến ước lượng khoảng cách mà không phụ thuộc vào cảm biến ngoại vi hay hạ tầng làn đường là mục tiêu trọng tâm mà đề tài này giải quyết.

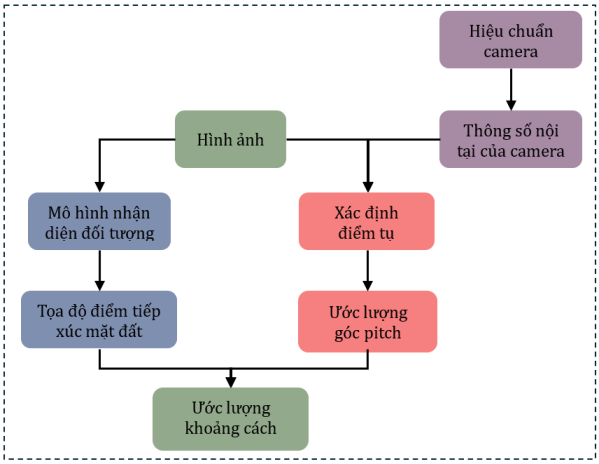
Nhóm phương pháp ước lượng khoảng cách bằng học sâu đang phát triển rất mạnh mẽ. Eigen et al. [12] đã đề xuất một trong những giải pháp tiêu biểu với kiến trúc mạng nơ-ron đa tầng: một tầng thực hiện dự đoán tổng quát dựa trên toàn bộ hình ảnh và tầng còn lại tinh chỉnh dự đoán ở mức cục bộ. Bằng cách sử dụng các bộ dữ liệu thô như NYU Depth và KITTI, phương pháp này có thể xác định chính xác các ranh giới chi tiết trong ảnh mà không cần chia cắt hình ảnh thành các vùng nhỏ. Nhằm khắc phục nhược điểm về nhu cầu dữ liệu gán nhãn thủ công khổng lồ [13] một số nghiên cứu đã hướng tới việc huấn luyện mạng nơ-ron theo cơ chế không giám sát, sử dụng dữ liệu từ camera stereo hoặc chuỗi video không có cấu trúc để tự học các đặc trưng về khoảng cách [14]. Mặc dù các mô hình này có khả năng hiểu ngữ cảnh rất tốt, chúng thường yêu cầu tài nguyên phần cứng mạnh mẽ để xử lý, khiến việc đạt được sự cân bằng giữa độ chính xác và tốc độ thực thi trên các hệ thống nhúng chi phí thấp vẫn là một thách thức lớn.

Tổng hợp các hướng tiếp cận trên cho thấy những hạn chế nhất định của từng nhóm phương pháp. Nhằm hướng tới một giải pháp cân bằng, nghiên cứu này vận dụng một quy trình tích hợp dựa trên mạng YOLOv8s và mô hình camera lỗ kim. Quy trình bắt đầu bằng việc chuẩn hóa thông số nội tại thông qua phương pháp hiệu chuẩn của Zhang, tạo nền tảng vững chắc cho các phép tính hình học. Tiếp đó, chúng tôi tận dụng kỹ thuật trích xuất cấu trúc tuyến tính bằng thuật toán LSD và RANSAC để tự động xác định góc pitch cho ảnh đơn đầu vào. Sự kết hợp giữa hiệu chuẩn tĩnh và hiệu chuẩn động giúp hệ thống thích ứng tốt hơn với sự biến đổi góc nhìn mà không cần quy trình cài đặt ngoại vi phức tạp hay cảm biến IMU hỗ trợ. Hướng tiếp cận này tập trung vào tính thực tiễn, đảm bảo sự chính xác ổn định trong khi vẫn duy trì hiệu suất tính toán phù hợp cho các thiết bị nhúng.

III. PHƯƠNG PHÁP ĐỀ XUẤT

A. MÔ HÌNH TỔNG QUAN

Phương pháp đề xuất nhằm ước lượng khoảng cách vật thể từ ảnh camera đơn. Trong nghiên cứu này, chúng tôi tập trung nghiên cứu đến sự ảnh hưởng của góc pitch (β) – được định nghĩa là góc hợp bởi trục quang học của camera và mặt phẳng nằm ngang của mặt đường. Cấu trúc tổng thể của phương pháp được thể hiện trong Hình 1.



Hình 1. *Sơ đồ tổng quan của phương pháp đề xuất.*

Có 4 giai đoạn chính trong phương pháp đề xuất này, bao gồm: xác định thông số nội tại của camera bằng phương pháp của Zhang, nhận diện đối tượng và xác định tọa độ điểm tiếp xúc với mặt phẳng đường bằng YOLOv8s, xác định điểm tụ và góc pitch bằng thuận toán LSD và RANSAC, và xây dựng công thức ước lượng khoảng cách.

B. DỮ LIỆU ĐẦU VÀO

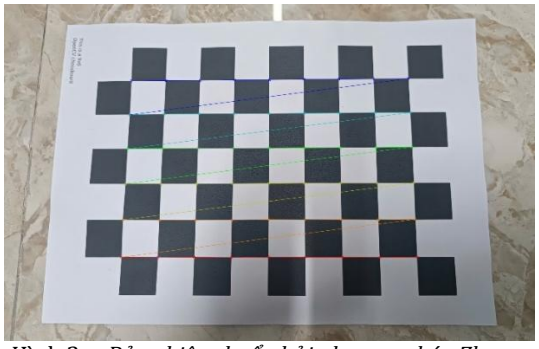
Trong nghiên cứu này, chúng tôi sử dụng hai bộ dữ liệu khác nhau là: bộ dữ liệu KITTI [15] và bộ dữ liệu tự thu thập. Bộ dữ liệu KITTI được ghi lại từ một hệ thống nhiều camera, cảm biến gắn trên xe và di chuyển xung quanh thành phố Karlsruhe, Đức. Nghiên cứu chỉ sử dụng các hình ảnh có đối tượng trong khoảng cách khoảng 25,00 m và độ cao camera có giá trị từ 1,60 m đến 1,70 m. Đối với dữ liệu tự thu thập, chúng tôi chụp ảnh người ở trong nhà và xe gắn máy trên mặt đường phẳng ngoài trời có khoảng cách đến camera từ 3,00 m đến 45,00 m.

C. HIỆU CHUẨN CAMERA

Hiệu chuẩn camera là bước cần thiết trong lĩnh vực thị giác máy tính nhằm trích xuất các thông tin nội tại của camera từ ảnh 2D. Tính linh hoạt, độ bền vững và chi phí thấp là những yếu tố quan trọng trong tác vụ hiệu chuẩn. Và phương pháp Zhang được phát triển dựa trên các tiêu chí đó. Phương pháp Zhang yêu cầu camera quan sát một mẫu phẳng ở các hướng khác nhau. Cả mẫu và camera có thể di chuyển tùy ý không cần biết trước. Phương pháp này là sự kết hợp giữa hai phương pháp trước đó với độ linh hoạt và ổn định tốt hơn nhiều [16]. Theo nghiên cứu của Zhang, để độ chính xác đạt kết quả tối ưu nhất thì số lượng ảnh chụp từ 13 đến 15 ảnh và góc chụp nằm trong khoảng giá trị từ 50° đến 70°.

Nghiên cứu này sử dụng bảng hiệu chuẩn như Hình 2 có kích thước 7x10 ô với bề rộng mỗi ô là 24,7cm x 24,7cm. Số lượng ảnh chụp là 15 ảnh ở những khoảng cách và góc nhìn khác nhau. Ảnh chụp sẽ được xử lý bằng thư viện của OpenCV chạy trên nền tảng Google Colab để tìm các thông số nội tại của camera. Đối với dữ liệu tự thu thập, chúng tôi sử dụng phương pháp Zhang để hiệu chuẩn camera và được kết quả sau:

$$R = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1906{,}56 & 0 & 1262{,}64 \\ 0 & 1903{,}46 & 963{,}17 \\ 0 & 0 & 1 \end{bmatrix}$$



Hình 2. *Bảng hiệu chuẩn bởi phương pháp Zhang.*

D. NHẬN DIỆN ĐỐI TƯỢNG

Nhận diện đối tượng là một trong những nhiệm vụ trọng yếu của thị giác máy tính. **Bảng 1** cho thấy trong số các mô hình tiên tiến hiện nay, dòng thuật toán YOLO vẫn được đánh giá cao nhờ sự tối ưu vượt trội giữa tốc độ xử lý và độ chính xác. Mặc dù xu hướng hiện nay đang chứng kiến sự trỗi dậy của các công nghệ mới dựa trên Transformer như RT-DETR, vốn thể hiện ưu thế trong những kịch bản đòi hỏi độ chính xác và tính ổn định cực cao như ảnh nhiệt hay môi trường phức tạp, nhưng YOLO vẫn giữ vai trò then chốt trong các ứng dụng yêu cầu phản hồi theo thời gian thực [17].

Bảng 1. *Kết quả các chỉ số đạt được trên bộ dữ liệu HORD của các mô hình YOLO và RT-DETRl.*

Models	Precision	Recall	mAP50	mAP50-95	Latency (ms)	Training time (hr)
YOLOv8l	0,91	0,827	0,918	0,703	14,7	2,43
YOLOv9c	0,86	0,854	0,911	0,684	12,4	2,76
YOLOv10l	0,801	0,689	0,811	0,611	13,9	3,7
YOLOv11l	0,883	0,822	0,889	0,671	12	2,14
RT-DETRl	0,801	0,742	0,783	0,55	14,9	3,67

Trong các mô hình của YOLO, phiên bản YOLOv8s đã khẳng định được vị thế là một kiến trúc đáng tin cậy, có sự hỗ trợ mạnh mẽ từ cộng đồng và hiệu suất ổn định, đặc biệt phù hợp để triển khai trên các hệ thống thực tế yêu cầu tốc độ xử lý nhanh [18].

Bảng 2. *So sánh hiệu quả của một số phiên bản YOLO mới nhất [19].*

Version	Params (M)	GFLOPs	mAP (50:95) on VOC07+12	mAP (50:95) on COCO2017
YOLOv7-S	37,3	105,4	71,2	51,4
YOLOv7-X	70,9	188,4	72,5	53,1
YOLOv8-S	11,1	28,7	67,6	44,9
YOLOv8-X	68,2	258,2	73,4	53,9
YOLOv9-E	70,0	245,5	76,0	55,6
YOLOv10-S	7,20	21,6	69,7	46,3
YOLOv10-X	29,5	160,4	74,0	54,4

Dữ liệu từ Bảng 2 cho thấy mặc dù các phiên bản như YOLOv8-X hay YOLOv9-E đạt mAP cao hơn, nhưng đi kèm với đó là yêu cầu lớn về số lượng tham số và sức mạnh tính toán (GFLOPs). Trong khi đó, các phiên bản mới nhất như YOLOv10S thể hiện hiệu năng tốt và nhẹ hơn nhưng chưa ổn định.

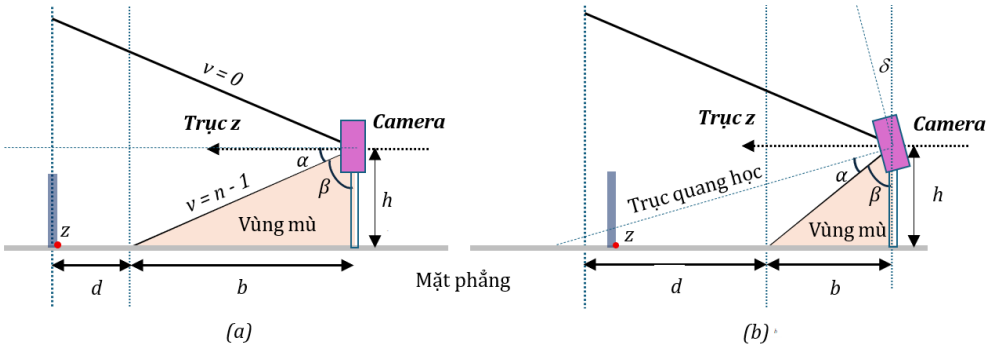
Trong khung nghiên cứu hướng tới các giải pháp ADAS chi phí thấp, việc lựa chọn mô hình phải đảm bảo sự cân bằng giữa độ chính xác và tài nguyên phần cứng. Phiên bản YOLOv8s cung cấp độ chính xác đủ tin cậy để xác định điểm tiếp xúc mặt đất, đồng thời đảm bảo khả năng vận hành mượt mà trên các thiết bị nhúng có tài nguyên hạn chế. Hiệu năng của mô hình sẽ được thể hiện trong phần kết quả thực nghiệm.

E. ƯỚC LƯỢNG KHOẢNG CÁCH

Mô hình camera lỗ kim mô tả mối quan hệ hình học giữa các điểm trong không gian 3D và hình chiếu của chúng trên mặt phẳng ảnh nên được sử dụng để ước lượng khoảng cách đến vật thể. Việc ước lượng khoảng cách thường được thực hiện theo hai cách: dựa vào thông tin của camera hoặc dựa vào kích thước của vật thể.

1. ƯỚC LƯỢNG KHOẢNG CÁCH DỰA VÀO TRƯỜNG NHÌN FOV

FOV là khu vực tối đa mà camera có thể thu hình. Nó liên quan tới tiêu cự của ống kính và kích thước cảm biến [20]. Việc triển khai mô hình FOV có thể được chia thành hai trường hợp là trục quang học của camera nằm ngang và trục có góc lệch tùy ý được mô phỏng trong Hình 3.



Hình 3. *Mô hình camera tương ứng với (a) không có góc pitch và (b) có góc pitch bất kỳ [20].*

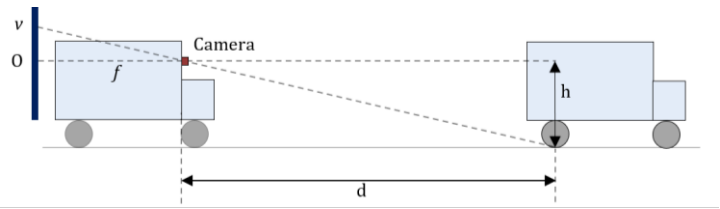
Công thức ước lượng khoảng cách theo phương pháp trường nhìn FOV:

$$z = h.tan\left[\beta + 2\alpha\left(\frac{n-1-v}{n-1}\right)\right]$$

2. PHƯƠNG PHÁP ĐỀ XUẤT

Dựa vào mô hình camera lỗ kim như Hình 4, ta có thể ước lượng khoảng cách theo công thức:

$$z = \frac{f.H}{v - v_0}$$



Hình 4. *Mô hình camera lỗ kim với trục quang học nằm ngang.*

Trường hợp trục quang học của camera lệch một góc β so với mặt phẳng nằm ngang như Hình 5. Nếu biết tọa độ vc của ảnh chiếu của điểm P lên mặt phẳng ảnh Πc, tương ứng với một camera có trục quang học lệch một góc β so với mặt phẳng ngang, ta có thể tính được tọa độ vv của ảnh chiếu của cùng điểm P đó lên mặt phẳng ảnh Πv, tương ứng với một camera ảo có hệ tọa độ không lệch so với mặt phẳng ngang [21]. Cụ thể, mối liên hệ như sau:

$$v_v = f.\frac{v^c.cos\beta - f.sin\beta}{v^c.sin\beta + f.cos\beta}$$

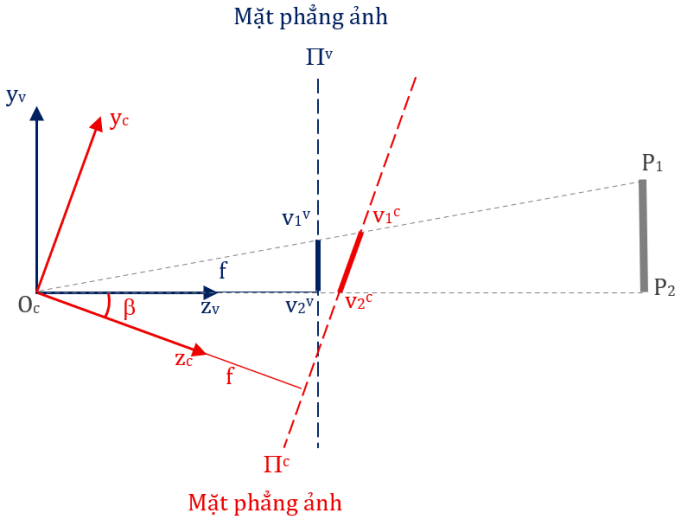
Biến đổi tọa độ để phù hợp với quy ước tọa độ đầu ra của YOLOv8s:

$$v_v - v_0 = f.\frac{(v^c - v_0).cos\beta - f.sin\beta}{(v^c - v_0).sin\beta + f.cos\beta}$$

Thay vào biểu thức, ta được:

$$z = \frac{f.H}{v_v - v_0} = H\frac{(v^c - v_0).sin\beta + f.cos\beta}{(v^c - v_0).cos\beta - f.sin\beta}$$

Trong đó, f là tiêu cự của camera, H là chiều cao camera so với mặt phẳng đường, v^c là tọa độ điểm của vật nằm trên mặt phẳng đường và v_0 là tọa độ tâm quang học theo trục y.



Hình 5. *Phép chiếu phối cảnh của một đoạn thẳng trên hai mặt phẳng ảnh với các hướng đặt camera khác nhau.*

3. ƯỚC LƯỢNG KHOẢNG CÁCH DỰA VÀO KÍCH THƯỚC VẬT THỂ

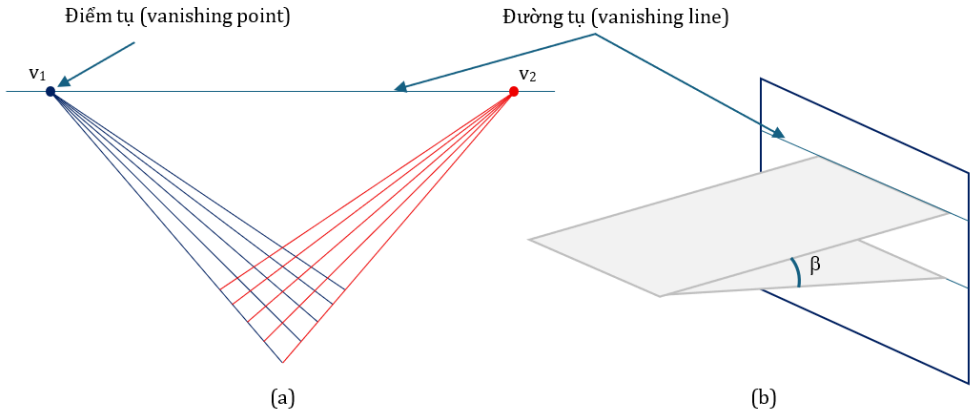
Sử dụng kích thước vật thể (chiều cao hoặc chiều rộng) cũng được sử dụng để ước lượng khoảng cách. Tuy nhiên, kích thước của các vật thể có sự thay đổi đáng kể nên độ chính xác của phương pháp này tương đối thấp. Do đó, phương pháp này có thể sử dụng như một phép kiểm tra.

$$z = \frac{f.H}{h}$$

Trong đó, f là tiêu cự của camera, H và h lần lượt là độ cao thực tế và độ cao trong ảnh của vật thể.

F. XÁC ĐỊNH GÓC HỢP BỞI TRỤC QUANG HỌC CỦA CAMERA VÀ MẶT PHẲNG ĐƯỜNG

Trong phép chiếu xuyên tâm, hình ảnh của một vật thể kéo dài đến vô tận có thể xuất hiện dưới dạng một thực thể hữu hạn trong ảnh. Điểm tụ (vanishing point) là hình ảnh của một điểm ở vô cực, đại diện cho hướng của một tập hợp các đường thẳng song song trong không gian 3D. Do đó, các đường thẳng song song trong không gian nhưng không song song với mặt phẳng ảnh sẽ hội tụ tại cùng một điểm tụ trong ảnh [22]. Trong cùng một mặt phẳng, các đường thẳng không song song với nhau sẽ có những điểm tụ khác nhau. Và tập hợp các điểm tụ khác nhau đó sẽ tạo thành đường tụ (vanishing line) được biểu diễn trong Hình 6. Đường tụ chỉ phụ thuộc vào hướng của mặt phẳng, không phụ thuộc vào vị trí của nó trong không gian.

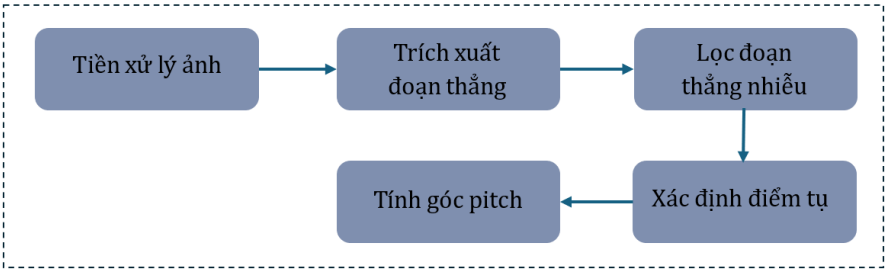


Hình 6. *Các đường thẳng song song với nhau sẽ hội tụ tại một điểm tụ (a), và các đường thẳng trên hai mặt phẳng không song song với nhau sẽ có đường tụ khác nhau (b).*

Khi trục quang học của camera song song với mặt phẳng nằm ngang, thì các đường thẳng song song với mặt phẳng đó sẽ có đường tụ đi qua tâm ảnh có tọa độ (u_0, v_0). Và khi thay đổi góc pitch thì các đường song song với mặt phẳng nằm ngang sẽ có đường tụ không đi qua tâm quang học.

Góc pitch được tính theo công thức:

$$\beta = arctan\left(\frac{v^c - v_0}{f}\right)$$



Hình 7. *Sơ đồ thể hiện các bước trong giai đoạn ước lượng góc pitch.*

Giai đoạn ước lượng góc pitch thể hiện trong Hình 7. bắt đầu bằng bước tiền xử lý ảnh nhằm chuẩn bị dữ liệu phù hợp cho các thuật toán hình học tiếp theo. Ảnh đầu vào trước hết được chuyển sang thang xám (grayscale) để giảm độ phức tạp tính toán nhưng vẫn giữ được thông tin cấu trúc cần thiết. Sau đó, ảnh được lọc nhiễu bằng bộ lọc bilateral. Phương pháp này giúp làm mịn các vùng đồng nhất trong ảnh trong khi vẫn bảo toàn biên, nhờ đó duy trì tốt các cấu trúc hình học quan trọng phục vụ cho việc phát hiện các đoạn thẳng. Trên ảnh đã được tiền xử lý, kỹ thuật phát hiện đoạn thẳng LSD (Line Segment Detector) được sử dụng để trích xuất các đoạn thẳng với độ chính xác dưới mức điểm ảnh [23]. Đây là một bước quan trọng nhằm thu được tập hợp các cấu trúc tuyến tính đặc trưng của cảnh, đặc biệt là các cạnh có xu hướng hội tụ về điểm tụ trong không gian ảnh. Tuy nhiên, kết quả từ LSD thường bao gồm cả những đoạn thẳng ngắn hoặc kém ổn định, có thể gây nhiễu cho quá trình ước lượng hình học. Vì vậy, cần thực hiện bước lọc để loại bỏ các đoạn thẳng có độ dài nhỏ hơn một ngưỡng nhất định. Việc này giúp giảm số lượng dữ liệu nhiễu và tăng độ ổn định khi xác định điểm tụ. Sau khi thu được tập hợp các đoạn thẳng đã được lọc, thuật toán RANSAC được áp dụng để tìm điểm tụ một cách bền vững. Cụ thể, thuật toán lặp lại quá trình chọn ngẫu nhiên hai đoạn thẳng để tính giao điểm tạm thời, sau đó đánh giá mức độ phù hợp của giao điểm này dựa trên số lượng các đoạn thẳng khác có hướng đi về gần điểm đó trong phạm vi sai số cho phép là 5 pixel. Quá trình này được lặp lại 1500 lần nhằm tăng khả năng tìm được nghiệm tốt nhất, ngay cả khi dữ liệu chứa nhiều nhiễu. Cuối cùng, góc pitch được suy ra từ vị trí điểm tụ thu được trên ảnh.

IV. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

A. MÔI TRƯỜNG THỰC NGHIỆM

Nghiên cứu này tập trung vào sự ảnh hưởng của góc pitch đến ước lượng khoảng cách và các đối tượng cần quan tâm nằm ở vị trí ngay phía trước camera. Vì vậy, nghiên cứu xem như góc lệch theo hai phương còn lại là không đáng kể và bỏ qua ảnh hưởng của chúng đến kết quả thực nghiệm. Ảnh dữ liệu đầu vào được chụp trong điều kiện ánh sáng tốt và không bị nhòe do chuyển động. Tuy nhiên, nghiên cứu có tạo thêm một bộ ảnh mô phỏng ảnh bị nhòe do chuyển động rung rắc để phân tích hiệu quả của mô hình.

B. KẾT QUẢ THỰC NGHIỆM

1. NHẬN DIỆN ĐỐI TƯỢNG

Chúng tôi tiến hành đánh giá một số mô hình phát hiện đối tượng tiêu biểu gồm YOLOv5s, YOLOv8s, YOLOv8m, YOLOv10s và RT-DETR trên bộ dữ liệu KITTI đã chuẩn bị nhằm khảo sát đồng thời hiệu năng tính toán và sai số hình học theo khoảng cách. Các mô hình được sử dụng với trọng số huấn luyện sẵn, không thực hiện tinh chỉnh nhằm đảm bảo tính khách quan trong quá trình so sánh. Hiệu năng được đánh giá dựa trên sai số (pixel) vị trí đáy của điểm tiếp xúc của vật thể với mặt phẳng đất theo từng khoảng cách.

Bảng 3. *Kết quả so sánh sai số điểm tiếp xúc các mô hình.*

Mô hình	0-10m	10-20m	20-30m	30m+	Latency_ms	FPS	Params_M
YOLOv5s	7.13	2.71	2.75	1.91	11.31	88.39	9.15
YOLOv8s	6.89	1.84	2.46	1.63	13.11	76.26	11.17
YOLOv8m	7.00	2.32	2.37	1.10	19.75	50.64	25.90
YOLOv10s	7.20	2.37	2.07	0.99	15.15	66.01	8.13
RT-DETR	6.45	2.30	1.37	1.46	50.64	19.75	32.97

2. ƯỚC LƯỢNG KHOẢNG CÁCH

Trong phần trình bày kết quả thực nghiệm, kí hiệu d là khoảng cách thực tế, Δd là sai số tuyệt đối, σ là sai số tương đối. Trong khi đó, các kí hiệu d_0, d_1 , d_2, d_3 lần lượt là khoảng cách ước lượng ứng với các điều kiện khác nhau.

Để tiếp cận nhanh và có thể so sánh với các phương pháp khác, chúng tôi đã tiến hành thực nghiệm trên bộ dữ liệu KITTI. Các ảnh đầu vào từ bộ dữ liệu này được lựa chọn để phù hợp với điều kiện camera có chiều cao không đổi và góc pitch bằng không. Ngoài ra, từ thông số nội tại camera được cung cấp trước, phương pháp FOV cũng được áp dụng để ước lượng khoảng cách.

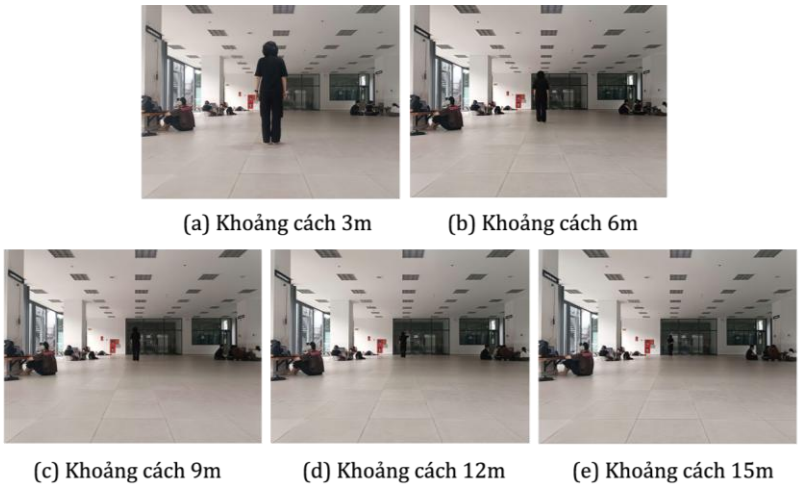
Bảng 4. *Kết quả thực nghiệm ước lượng khoảng cách và sai số trên bộ dữ liệu KITTI.*

STT	Hình ảnh	d	FOV			Phương pháp đề xuất		
			$d_{ước\ lượng}$ (m)	Δd (m)	σ (%)	$d_{ước\ lượng}$ (m)	Δd (m)	σ (%)
1	000125.png	7,01	7,50	0,49	6,99	7,06	0,05	0,67
2	000513.png	7,76	8,17	0,41	5,26	7,63	0,13	1,68
3	000468.png	8,49	9,75	1,26	14,81	8,87	0,38	4,45
4	000268.png	9,15	9,95	0,80	8,72	9,03	0,12	1,34
5	000409.png	9,29	9,95	0,66	7,08	9,03	0,26	2,83
6	000036.png	10,13	10,62	0,49	4,80	9,57	0,56	5,49
7	000275.png	13,14	15,94	2,80	21,30	13,58	0,44	3,36
8	000522.png	13,50	14,92	1,42	10,55	12,85	0,65	4,78
9	000462.png	14,65	17,32	2,67	18,26	14,55	0,10	0,65
10	000101.png	19,47	23,94	4,47	22,95	18,81	0,66	3,39
11	000292.png	19,65	25,28	5,63	28,63	19,60	0,05	0,26
12	000285.png	20,13	23,22	3,09	15,37	18,42	1,71	8,51
13	000060.png	21,02	26,71	5,69	27,07	20,62	0,40	1,90
14	000374.png	23,04	29,90	6,86	29,79	22,24	0,80	3,45
15	000007.png	25,01	34,72	9,71	38,84	24,78	0,23	0,91
16	000080.png	29,39	46,39	17,00	57,85	30,40	1,01	3,43

Bảng 5. *So sánh sai số trung bình tuyệt đối của các phương pháp trên bộ dữ liệu KITTI*

Phương pháp	MAE (m)		
	0-9m	10-19m	20-29m
XGBoost	0,377	0,603	0,975
LSTM	0,415	0,525	0,869
RF	0,408	0,629	1,059
FOV	0,724	2,370	7,997
Phương pháp đề xuất	0,188	0,482	0,700

Để kiểm tra hiệu quả của toàn bộ phương pháp, chúng tôi tiến hành sử dụng ảnh tự thu thập được chụp bởi camera không biết trước tham số nội tại. Ảnh được chụp trong điều kiện góc pitch bằng không và không thay đổi như Hình 8.



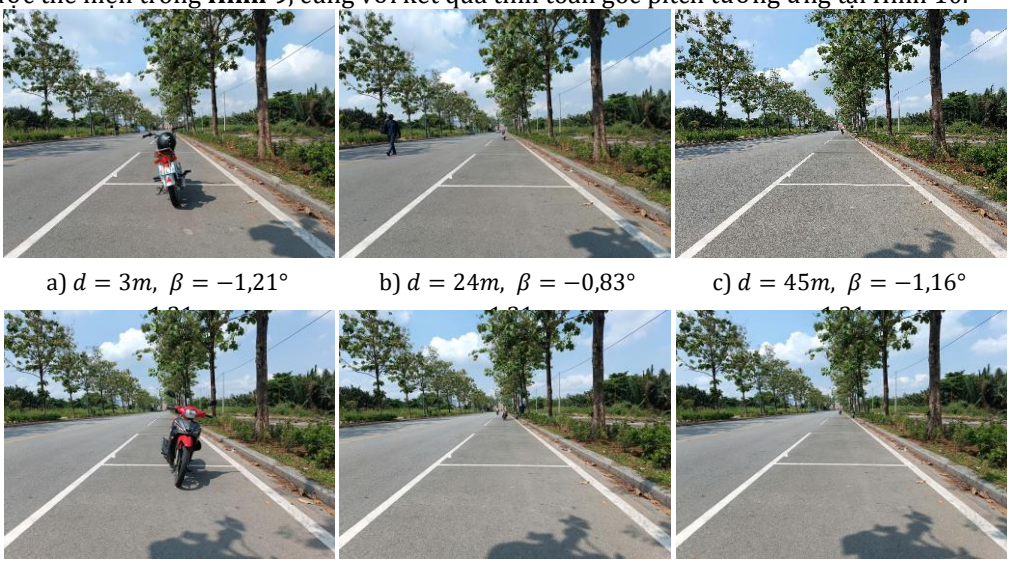
Hình 8. *Dữ liệu thực nghiệm trong nhà với camera cố định nằm ngang so với mặt phẳng ngang*

Ngoài ra, thực nghiệm thay đổi các thông số ban đầu như hạt giống ngẫu nhiên (random seed) và số lần lặp (iterations) trong thuật toán RANSAC để kiểm tra sự ảnh hưởng của chúng đến kết quả.

Bảng 6. *Kết quả thực nghiệm ước lượng khoảng cách trên bộ dữ liệu trong nhà.*

Ảnh	d (m)	Không áp dụng góc pitch (β)			Có áp dụng góc pitch (β)					
		d_0 (m)	Δd (m)	σ (%)	d_1 (m)	d_2 (m)	d_3 (m)	$\bar{d}$ (m)	$\Delta\bar{d}$ (m)	σ (%)
1	3	3,040	0,040	1,33	3,057	3,065	3,070	3,064	0,064	2,13
2	6	6,082	0,082	1,37	6,130	6,130	6,111	6,124	0,124	2,06
3	9	8,815	0,185	2,06	8,917	8,917	8,937	8,924	0,076	0,85
4	12	11,574	0,426	3,55	11,338	11,679	12,004	11,674	0,326	2,72
5	15	14,409	0,591	3,94	14,096	15,500	14,681	14,759	0,241	1,61

Bộ dữ liệu ngoài trời bao gồm 15 ảnh có góc pitch khoảng $-1°$ và 15 ảnh có góc pitch khoảng $-5°$. Đối tượng cần xác định là xe máy có khoảng cách từ 3,00 m đến 45,00 m. Một số mẫu ảnh với các biến thiên về khoảng cách và góc pitch được thể hiện trong **Hình 9**, cùng với kết quả tính toán góc pitch tương ứng tại Hình 10.



Hình 9. *Dữ liệu thực nghiệm ngoài trời có khoảng cách và góc pitch thay đổi.*



a) $d = 3m,\ \beta = -1{,}21°$ b) $d = 24m,\ \beta = -0{,}83°$ c) $d = 45m,\ \beta = -1{,}16°$

-1.21 -1.21 -1.21

Hình 10. *Ảnh sau khi áp dụng kỹ thuật LSD và thuật toán RANSAC.*

Kết quả thực nghiệm với bộ dữ liệu ngoài trời được thể hiện trong Bảng 7.

Bảng 7. *Kết quả thực nghiệm ước lượng khoảng cách trên bộ dữ liệu ngoài trời.*

d (m)	β (°)	$d_{ước\ lượng}$ (m)	Δd (m)	σ (%)	β (°)	$d_{ước\ lượng}$ (m)	Δd (m)	σ (%)
3	-1,21	3,11	0,11	3,52	-5,73	3,10	0,10	3,50
6	-1,04	6,24	0,24	3,96	-5,75	6,13	0,13	2,20
9	-0,93	9,33	0,33	3,66	-5,73	9,33	0,33	3,69
12	-0,79	12,76	0,76	6,31	-5,69	12,67	0,67	5,62
15	-0,80	15,54	0,54	3,59	-5,53	16,17	1,17	7,79
18	-1,00	19,01	1,01	5,62	-5,76	18,78	0,78	4,34
21	-0,89	21,52	0,52	2,49	-5,75	21,87	0,87	4,16
24	-0,83	25,57	1,57	6,55	-5,70	25,26	1,26	5,23
27	-0,78	28,09	1,09	4,03	-5,76	27,04	0,04	0,17
30	-0,84	30,82	0,82	2,74	-5,59	31,34	1,34	4,46
33	-1,26	33,57	0,57	1,72	-5,60	33,94	0,94	2,84
36	-1,39	37,31	1,31	3,63	-5,69	38,75	2,75	7,63
39	-1,20	39,08	0,08	0,21	-5,78	43,11	4,11	10,53
42	-1,03	43,21	1,21	2,87	-5,95	39,81	-2,19	-5,23
45	-1,16	41,95	-3,05	-6,78	-5,96	43,09	-1,91	-4,25

3. ẢNH HƯỞNG CỦA CHUYỂN ĐỘNG RUNG LẮC

Để khảo sát ảnh hưởng của rung lắc động đến việc trích xuất các đoạn thẳng qua kỹ thuật LSD từ đó dẫn đến sai lệch kết quả ước lượng khoảng cách, bộ dữ liệu thực nghiệm được tạo thêm các mức nhòe chuyển động theo phương dọc từ ảnh gốc. Cụ thể, ảnh được làm mờ bằng phép tích chập với một kernel nhòe tuyến tính có kích thước $L \times L$ theo phương thẳng đứng. **Hình 11** thể hiện ảnh có khoảng cách 3,00m được mô phỏng nhòe ở 4 mức độ khác nhau.



a) L = 15 pixel b) L = 20 pixel c) L = 25 pixel d) L = 30 pixel

-1.21 -1.21 -1.21

Hình 11. *Dữ liệu thực nghiệm mô phỏng nhòe chuyển động.*

Với bộ dữ liệu mô phỏng nhòe do chuyển động rung lắc, khoảng cách ước lượng ảnh hưởng bởi góc pitch và giai đoạn nhận diện đối tượng. Vì thế, các trường hợp mô hình không nhận diện được đối tượng sẽ có kết quả để trống trong Bảng 8.

Bảng 8. *Kết quả thực nghiệm ước lượng khoảng cách trong điều kiện mô phỏng nhòe chuyển động*

d (m)	**L = 0**		**L = 15**		**L = 20**		**L = 25**		**L = 30**	
	d_0 (m)	σ (%)	d_{15} (m)	σ (%)	d_{20} (m)	σ (%)	d_{25} (m)	σ (%)	d_{30} (m)	σ (%)
3	3,11	3,52	3,06	2,14	3,08	2,75	3,03	1,08	3,08	2,78
6	6,24	3,96	6,18	3,02	6,09	1,45	6,20	3,34	5,98	-0,32
9	9,33	3,66	9,16	1,81	-	-	8,90	-1,06	10,17	12,95
12	12,76	6,31	12,48	3,98	12,42	3,53	12,45	3,75	9,11	-24,06
15	15,54	3,59	15,54	3,60	15,13	0,85	-7,11	-147,39	13,26	-11,59
18	19,01	5,62	18,28	1,56	18,05	0,27	17,82	-0,98	-37,72	-309,55
21	21,52	2,49	20,52	-2,30	19,21	-8,55	20,51	-2,31	-23,10	-209,98
24	25,57	6,55	23,37	-2,60	22,91	-4,56	21,69	-9,63	18,71	-22,06
27	28,09	4,03	25,02	-7,35	24,17	-10,48	-285,53	-1157,53	-4,40	-116,30
30	30,82	2,74	27,03	-9,90	26,53	-11,57	28,51	-4,97	-	-
33	33,57	1,72	30,36	-8,01	27,57	-16,44	28,53	-13,53	-5,16	-115,63
36	37,31	3,63	30,70	-14,73	35,01	-2,74	31,71	-11,91	-	-
39	39,08	0,21	-	-	-	-	-	-	-	-
42	43,21	2,87	36,11	-14,03	-	-	-	-	-	-
45	41,95	-6,78	34,59	-23,13	-	-	-	-	-	-

C. ĐÁNH GIÁ KẾT QUẢ THỰC NGHIỆM

Kết quả từ **Bảng 3** cho thấy YOLOv8s đạt được sự cân bằng và ổn định tốt nhất giữa độ chính xác theo khoảng cách, số lượng tham số và tốc độ suy luận. Vì vậy, YOLOv8s là lựa chọn phù hợp để làm mô hình nhận diện đối tượng trong phương pháp đề xuất.

Với bộ dữ liệu KITTI, kết quả thực nghiệm từ Bảng 4 cho thấy phương pháp đề xuất có độ chính xác vượt trội và tính ổn định cao hơn hẳn so với phương pháp dựa trên trường nhìn FOV. Trong khi sai số của FOV tăng mạnh theo khoảng cách, lên đến 57,85% tại mốc gần 30 mét, thì phương pháp đề xuất của nghiên cứu vẫn duy trì sai số ở mức thấp dưới 9%. Kết quả Bảng 5 cho thấy khi so sánh với các thuật toán học máy và học sâu như Random Forest, XGBoost và LSTM, giải pháp đề xuất tiếp tục khẳng định hiệu quả cao với chỉ số sai số tuyệt đối trung bình (MAE) chỉ đạt 0,700m trong khoảng cách từ 20 đến 29 mét. Tuy nhiên, cũng cần lưu ý rằng kết quả so sánh này đang dựa trên các điều kiện thực nghiệm thuận lợi.

Với bộ dữ liệu tự thu thập, kết quả từ Bảng 6 đã chứng minh tính bền vững của hệ thống trước các yếu tố ngẫu nhiên. Các thử nghiệm thay đổi hạt giống (random seed) và số lần lặp trong thuật toán RANSAC cho thấy kết quả có sự chênh lệch không đáng kể. Tuy nhiên, trong điều kiện góc pitch chênh lệch đáng kể so với giả định thì sai số sẽ rất lớn. Kết quả thực nghiệm được thể hiện trong Bảng 7 cho thấy độ ổn định cao trong giai đoạn xác định góc pitch. Từ đó, sai số ước lượng khoảng cách tối đa là 10,53% trong phạm vi 45,00 m. Điều này khẳng định vai trò quan trọng của việc tích hợp góc pitch vào mô hình tính toán.

Trong điều kiện thực tế, ảnh hưởng bởi chuyển động rung lắc hoặc camera di chuyển với tốc độ cao sẽ ảnh hưởng đáng kể đến khả năng xử lý của mô hình. Trong phạm vị nghiên cứu, chúng tôi chỉ xét sự ảnh hưởng của chuyển động rung lắc theo phương dọc. Kết quả từ Bảng 8 cho thấy kỹ thuật LSD có hiệu quả ổn định khi ảnh bị nhiễu bởi kernel có độ dài L dưới 15 pixel. Khi L tăng lên 20 pixel, độ ổn định đến từ LSD vẫn tốt nhưng quá trình nhận diện đối tượng bắt đầu không đảm bảo hiệu quả. Trong giới hạn này thì khoảng cách ước lượng vẫn ổn định. Và khi L lớn hơn 25 pixel thì kỹ thuật LSD không còn ổn định và các đường được trích xuất bắt đầu có hiện tượng gãy, không còn hội tụ về điểm tụ dẫn đến ước lượng khoảng cách bị sai lệch rất lớn.

V. KẾT LUẬN

Nghiên cứu đã xây dựng thành công quy trình ước lượng khoảng cách dựa trên ảnh camera đơn bằng cách kết hợp học sâu và mô hình camera lỗ kim. Đóng góp cốt lõi của công trình là giải quyết được sự ảnh hưởng của góc pitch bằng kỹ thuật xác định điểm tụ qua thuật toán LSD và RANSAC. Kết quả thực nghiệm khẳng định tính hiệu quả của mô hình với sai số trung bình đạt mức 10,53% trong phạm vi 45,00 m, mở ra triển vọng ứng dụng thực tiễn trong việc tối ưu hóa các hệ thống giám sát giao thông và xe tự hành.

Mặc dù đạt được những kết quả khả quan, nghiên cứu vẫn tồn tại những giới hạn nhất định cần nhìn nhận khách quan. Kỹ thuật xác định điểm tụ có thể gặp nhiễu trong môi trường thiếu cấu trúc tuyến tính rõ ràng hoặc chịu tác

động từ sự rung lắc của phương tiện khi vận hành ở tốc độ cao. Bên cạnh đó, việc giả định mặt đường phẳng tuyệt đối và độ cao camera cố định có thể dẫn đến sai số khi di chuyển trên các địa hình phức tạp, lồi lõm.

Để tối ưu hóa phương pháp hướng tới ứng dụng thực tế, các nghiên cứu tiếp theo cần tập trung vào việc tăng cường khả năng trích xuất đặc trưng hình học trong điều kiện ánh sáng yếu hoặc môi trường có độ nhiễu cao. Đây sẽ là tiền đề quan trọng để phát triển các giải pháp thị giác máy tính chi phí thấp nhưng có độ tin cậy cao, phục vụ cho hệ thống hỗ trợ lái xe nâng cao (ADAS) và các công nghệ tự hành trong tương lai.

VI. LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Nguồn ngân sách khoa học và công nghệ Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong đề tài sinh viên nghiên cứu khoa học năm học 2025 – 2026.

VII. TÀI LIỆU THAM KHẢO

[1] H. Liao *et al.* (2024), "Real-time accident anticipation for autonomous driving through monocular depth-enhanced 3d modeling," *Accident Analysis & Prevention,* vol. 207, p. 107760.
[2] S. Lee, K. Han, S. Park, and X. Yang (2022), "Vehicle distance estimation from a monocular camera for advanced driver assistance systems," *Symmetry,* vol. 14, no. 12, p. 2657.
[3] X. Dong, M. A. Garratt, S. G. Anavatti, and H. A. Abbass (2022), "Towards real-time monocular depth estimation for robotics: A survey," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 10, pp. 16940-16961.
[4] D. Forouher, M. G. Besselmann, and E. Maehle (2016), "Sensor fusion of depth camera and ultrasound data for obstacle detection and robot navigation," in *2016 14th international conference on control, automation, robotics and vision (ICARCV)*, IEEE, pp. 1-6.
[5] Y. Zhou, Y. Dong, F. Hou, and J. Wu (2022), "Review on millimeter-wave radar and camera fusion technology," *Sustainability,* vol. 14, no. 9, p. 5114.
[6] R. Vernimmen, A. Hooijer, and M. Pronk (2020), "New ICESat-2 satellite LiDAR data allow first global lowland DTM suitable for accurate coastal flood risk assessment," *Remote Sensing,* vol. 12, no. 17, p. 2827.
[7] C. Zhao, Q. Sun, C. Zhang, Y. Tang, and F. Qian (2020), "Monocular depth estimation based on deep learning: An overview," *Science China Technological Sciences,* vol. 63, no. 9, pp. 1612-1627.
[8] D. Scharstein and R. Szeliski (2002), "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision,* vol. 47, no. 1, pp. 7-42.
[9] G. P. Stein, O. Mano, and A. Shashua (2003), "Vision-based ACC with a single camera: bounds on range and range rate accuracy," in *IEEE IV2003 intelligent vehicles symposium. Proceedings (Cat. No. 03TH8683)*, IEEE, pp. 120-125.
[10] K.-Y. Park and S.-Y. Hwang (2014), "Robust range estimation with a monocular camera for vision-based forward collision warning system," *The Scientific World Journal,* vol. 2014, no. 1, p. 923632.
[11] Y.-C. Chen, T.-F. Su, and S.-H. Lai (2014), "Integrated vehicle and lane detection with distance estimation," in *Asian Conference on Computer Vision*, Springer, pp. 473-485.
[12] D. Eigen, C. Puhrsch, and R. Fergus (2014), "Depth map prediction from a single image using a multi-scale deep network," *Advances in neural information processing systems,* vol. 27.
[13] R. Garg, V. K. Bg, G. Carneiro, and I. Reid (2016), "Unsupervised cnn for single view depth estimation: Geometry to the rescue," in *European conference on computer vision*, Springer, pp. 740-756.
[14] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe (2017), "Unsupervised Learning of Depth and Ego-Motion from Video," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21-26 July, pp. 6612-6619, doi: 10.1109/CVPR.2017.700.
[15] A. Geiger, P. Lenz, and R. Urtasun (2012), "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, pp. 3354-3361.
[16] Z. Zhang (2000), "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 11, pp. 1330-1334.
[17] A. J. G. Hernández, J. P. S. Hernández, D. L. H. Rabadán, J. F. Solis, and J. G. Barbosa (2025), "An analysis of YOLO models versus RT-DETR applied to multi-object detection in images," *International Journal of Combinatorial Optimization Problems and Informatics,* vol. 16, no. 3, p. 360.
[18] M. Sohan, T. Sai Ram, and C. V. Rami Reddy (2024), "A review on yolov8 and its advancements," in *International conference on data intelligence and cognitive informatics*, Springer, pp. 529-545.
[19] Z. Hua *et al.* (2025), "A benchmark review of YOLO algorithm developments for object detection," *IEEE Access.*
[20] K. S. Chou *et al.* (2023), "A lightweight robust distance estimation method for navigation aiding in unsupervised environment using monocular camera," *Applied Sciences,* vol. 13, no. 19, p. 11038.
[21] M. Momeni-k, S. C. Diamantas, F. Ruggiero, and B. Siciliano (2012), "Height estimation from a single camera view," in *International Conference on Computer Vision Theory and Applications*, vol. 2: SciTePress, pp. 358-364.

[22] R. Hartley and A. Zisserman (2003), *Multiple view geometry in computer vision*. Cambridge university press.
[23] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall (2010), "LSD: a fast line segment detector with a false detection control," *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, no. 4, pp. 722-732.

MONOCULAR DISTANCE ESTIMATION UNDER VARYING PITCH ANGLES USING DEEP LEARNING AND THE PINHOLE CAMERA MODEL

Tran Khac Tri, Lam Thi Hien, Bui Lien Tue Nhi, Nguyen Minh Hai*

ABSTRACT — This study addresses the problem of distance estimation to preceding vehicles in Advanced Driver Assistance Systems (ADAS) using a monocular camera. A fundamental challenge in traditional geometric models lies in the systematic errors caused by dynamic variations in the camera pitch angle under real-world operating conditions. We propose an integrated framework that combines the YOLOv8s deep learning network for object detection with a pinhole camera model for distance calculation. In this study, road geometric features are extracted using the Line Segment Detector (LSD) and a robust RANSAC algorithm to determine the vanishing point, enabling real-time pitch angle compensation without the need for costly IMU sensors. Experimental results validate the effectiveness of the proposed model, achieving an average error of 10.53% within a 45.00-meter range. This study not only ensures measurement reliability but also optimizes computational performance, providing a cost-effective solution for traffic collision warning systems and autonomous vehicles.

Keywords — camera calibration, object detection, pitch angle, distance estimation, computer vision



Trần Khắc Trí hiện đang là sinh viên khoa Vật Lý tại Trường Đại học Sư phạm Thành phố Hồ Chí Minh (HCMUE). Hướng nghiên cứu chính: ứng dụng AI để giải quyết các bài toán chuyên sâu vật lý hạt nhân.



Lâm Thị Hiền hiện đang là sinh viên khoa Vật Lý tại Trường Đại học Sư Phạm Thành phố Hồ Chí Minh (HCMUE). Hướng nghiên cứu chính: nghiên cứu về các phương pháp học máy.



Bùi Liên Tuệ Nhi hiện đang là sinh viên khoa Vật Lý tại Trường Đại học Sư Phạm Thành phố Hồ Chí Minh (HCMUE). Hướng nghiên cứu chính: ứng dụng AI để giải quyết các bài toán liên quan đến vật lý.



Nguyễn Minh Hải nhận bằng Tiến sĩ Khoa học máy tính tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (VAST) năm 2025. Hiện ông đang là giảng viên của Đại học Sư phạm Thành phố Hồ Chí Minh (HCMUE). Hướng nghiên cứu của ông tập trung vào trí tuệ nhân tạo, với định hướng phát triển và ứng dụng các mô hình học máy và học sâu trong các lĩnh vực y tế và giáo dục, nhằm nâng cao hiệu quả khai thác dữ liệu và giải quyết các vấn đề khoa học và công nghệ gắn với yêu cầu thực tiễn.