NHẬN DẠNG BIỂN BÁO GIAO THÔNG DÙNG MÁY HỌC

TRAFFIC SIGN RECOGNITION USING MACHINE LEARNING

Nguyễn Thụy Tô Hoài Nghi, Lê Mỹ Hà

Trường đại học Sư phạm Kỹ thuật TP.HCM

TÓM TẮT

Ngày nay, sự phát triển bùng nổ của Machine Learning/ Neural Network đã đưa máy móc đến gần với con người, đã có rất nhiều ứng dụng có kết quả khả quan trong đời sống, đặc biệt là giao thông. Trong bài báo này, tác giả sử dụng kỹ thuật đặc trưng kênh tổng hợp Aggregate Channel Features – ACF trong Machine Learning để nhận dạng đối tượng biển báo, xe máy và ô tô trong quá trình tham gia giao thông trên các tuyến đường khác nhau trong môi trường đô thị, bên cạnh đó tác giả đã kết hợp phát hiện làn đường bằng thuật toán mắt chim và dự đoán khoảng cách từ camera đơn đến vật thể để hỗ trợ cho việc cảnh báo va chạm phía trước. Kết quả cho thấy kỹ thuật ACF là một kỹ thuật hiệu quả trong việc nhận dạng với thuật toán đơn giản và tốc độ xử lý nhanh.

Từ khóa: *ACF; Aggregate Channel Features; đặc trưng kênh tổng hợp; nhận dạng đối tượng; kỹ thuật mắt chim.*

ABSTRACT

Today, the explosive growth of the Machine Learning / Neural Network has brought machines close to people, there are many applications that have positive results in our life, especially in the traffic. In this article, the author used Aggregate Channel Features (ACF) in Machine Learning to identify traffic sign, motorbikes and cars in roads. In addition, the author combined lane detection using the bird-eyes image algorithm and estimated the distance from mono camera to other objects to support frontal warning. The result showed that the ACF technique is an efficient technique with detecting simple algorithms and fast processing speeds.

Keywords: *ACF; Aggregate Channel Features; object detection; bird-eyes image technique.*

1. GIỚI THIỆU

Ngành AI – Artificial Intelligence (trí tuệ nhân tạo) đã được bắt đầu nghiên cứu từ thập niên 50, bắt đầu phát triển từ những năm 70, và phát triển một cách bùng nổ trong những năm gần đây, được nổi lên như cuộc cách mạng công nghiệp 4.0. Ngành AI hiện nay ứng dụng sâu rộng trong sản xuất cũng như trong đời sống, đặc biệt là các nước phát triển.

Một trong những ứng dụng rất quan trọng của AI đó là trong lĩnh vực giao thông. Tháng 4 năm 2017, tại hội thảo "Tương lai công nghệ ô tô và robot trong cuộc cách mạng công nghiệp 4.0" ở nước ta, FPT Software cũng giới thiệu mô hình thử nghiệm xe tự lái do công ty tự nghiên cứu phát triển từ tháng 8 năm 2016. Tại đây, các chuyên gia còn giới thiệu giải pháp công nghệ hỗ trợ lái xe tiên tiến (ADAS - Advanced Driver Assistance Systems) như nhận diện làn đường, camera cảm biến 360; phát hiện vật cản, tín hiệu đèn giao thông, biển báo giao thông, bảng giám sát điều khiển thân vỏ ô tô...

Khảo sát các nghiên cứu gần đây về vấn đề nhận dạng đối tượng thì chủ yếu xoay quanh vào việc tăng độ chính xác và tốc độ xử lý bằng cách sử dụng Deep Learning (một nhánh nổi trội của AI) kết hợp các thuật toán hay đưa ra một số thuật toán được sử dụng để nhận dạng như trong [1] sử dụng phương pháp đặc trưng kênh tổng hợp Aggregate Channel Features (ACF) – một kỹ thuật trong Machine Learning. ACF đã đưa được kết quả rất khả quan trong việc nhận dạng khuôn mặt người hay nhận dạng chữ viết như trong [2], [3].

Trong bài báo này, tác giả đã chọn ACF để thực hiện việc nhận dạng đối tượng biển báo, xe máy và ô tô trong môi trường đô thị. Ngoài

ra, tác giả kết hợp nhận dạng làn đường và dự đoán khoảng cách từ mono camera tới đối tượng được nhận dạng. Những vấn đề này chính là một trong những bài toán của hệ thống lái xe tự động đang phát triển mạnh mẽ hiện nay.

2. PHƯƠNG PHÁP THỰC HIỆN

Để thực hiện việc phát hiện làn đường và các đối tượng trên đường, tác giả sử dụng những video trên đường thực tế trong đô thị, lưu lại bằng camera điện thoại với cảm biến chính có độ phân giải 13MP, khẩu độ f/2.2 và kích thước cảm biến lớn 1/3inch, xử lý offline bằng phần mềm Matlab 2018a, với cấu hình máy tính như sau:

- System Information: Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz (4 CPUs), ~2.2GHz. Memory: 4096MB RAM.
- Display Devices: Card name: NVIDIA GeForce 940M. Memory: 3994 MB

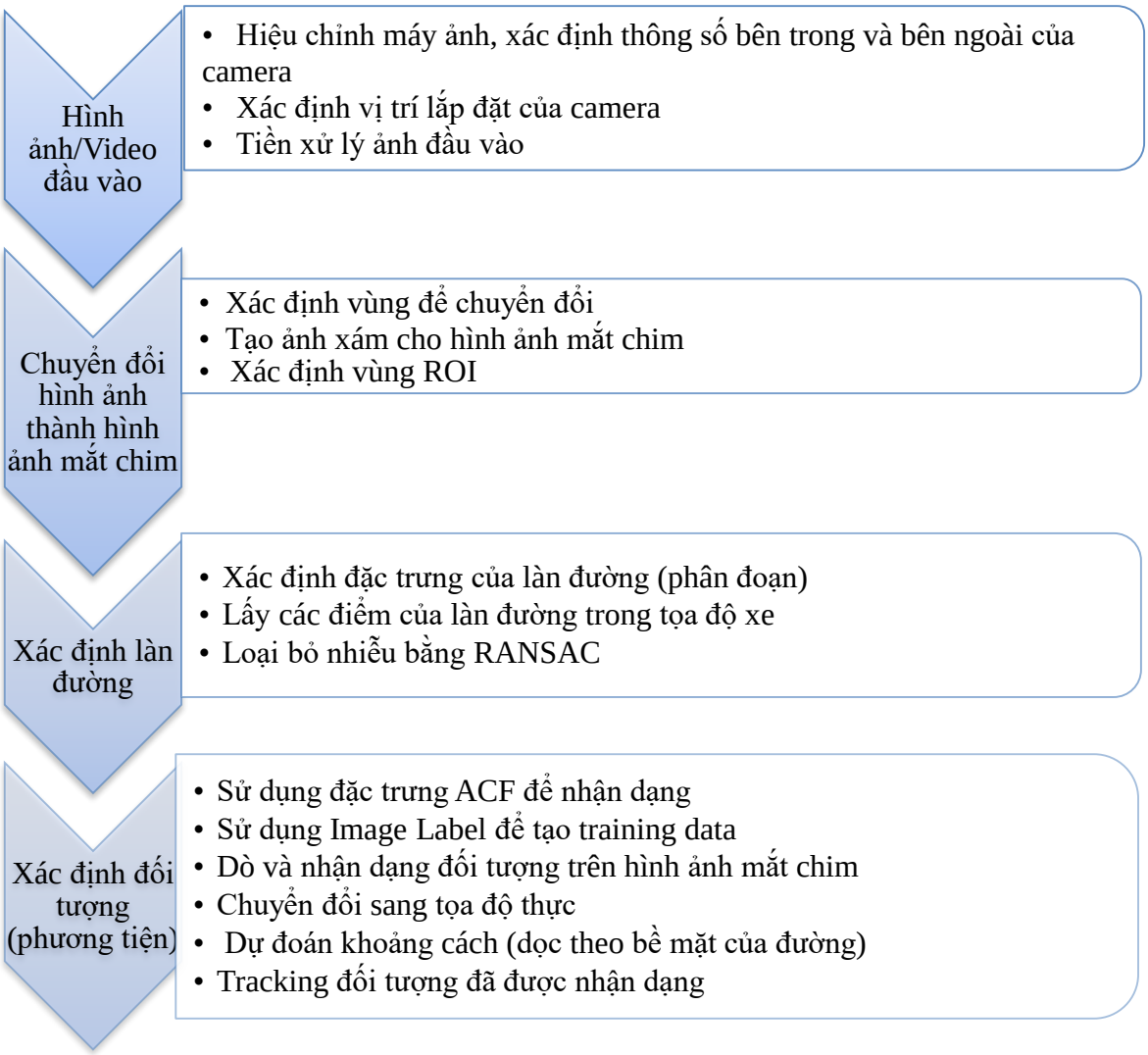
Qúa trình sử lý offline thực hiện theo hình 1.

2.1. Hiệu chỉnh máy ảnh

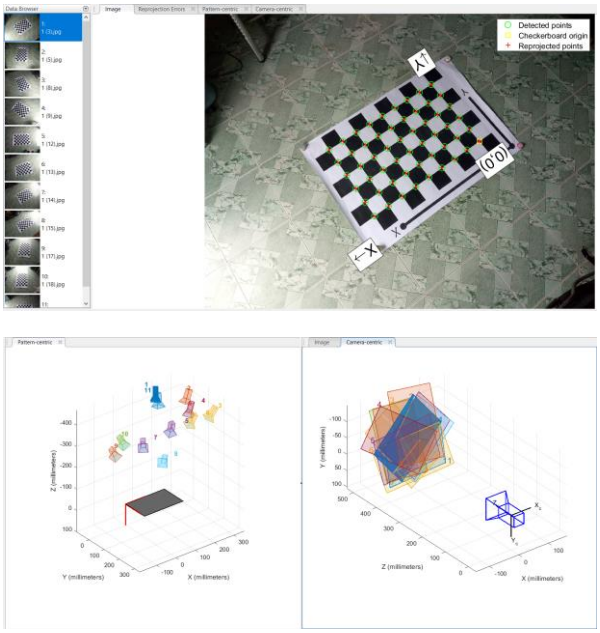
Những đối tượng trong hình ảnh được thu lại bằng camera đơn được nhận dạng chính xác hay việc dự đoán khoảng cách phụ thuộc rất nhiều vào các thông số camera sử dụng, bao gồm: thông số nội (bên trong), thông số ngoại (bên ngoài) và hệ số biến dạng. Để xác định các thông số này, tác giả sử dụng Camera Calibration toolbox của Matlab.

Để ước tính các thông số của camera đã sử dụng, tác giả sử dụng 11 bức ảnh chụp chess board, với khoảng cách từ camera tới vật thể trung bình là 2m.

Kết quả của quá trình này sẽ xác định được các thông số của camera để sử dụng cho các bước tính toán sau này.



Hình 1. *Lưu đồ thực hiện*



Hình 2. *Kết quả hiệu chỉnh camera*

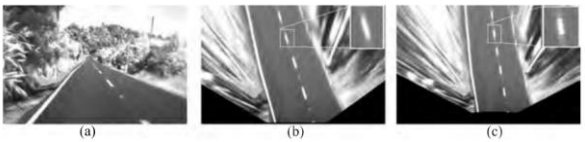
2.2. Phát hiện làn đường

2.2.1. Kỹ thuật mắt chim

Trong bài báo này, người thực hiện sử dụng phương pháp kỹ thuật mắt chim đảo tầm nhìn hình ảnh Inverse perspective mapping – IPM để phát hiện làn đường [4]. Kỹ thuật này đòi hỏi yêu cầu cao cho việc kiểm tra các thông số của camrera đầu vào, để đảm bảo cho việc chuyển đổi hình ảnh đạt được chất lượng cao hơn so với kĩ thuật uốn cong.

Theo phương pháp toán học, IPM có thể được mô hình hóa như một phép chiếu từ một không gian 3D Euclide W, chứa các phần tử $(x, y, z) \in R^2$ thành một không gian con phẳng 2D của R , được biểu hiễn bởi I, với các thành phần $(u, v) \in R^2$. Ánh xạ từ I sang W được tính như công thức (1):

$$\begin{aligned} x(u,v) &= h \cdot \cot\left\{(\theta_0 - \alpha_u) + u\frac{2\alpha_u}{m-1}\right\} \cdot \cos\left\{(\gamma_0 - \alpha_v) + v\frac{2\alpha_v}{n-1} + l\right\} \\ y(u,v) &= 0 \\ z(u,v) &= h \cdot \cot\left\{(\theta_0 - \alpha_u) + u\frac{2\alpha_u}{m-1}\right\} \cdot \cos\left\{(\gamma_0 - \alpha_v) + v\frac{2\alpha_v}{n-1} + d\right\} \end{aligned} \quad (1)$$



Hình 3. *Hình ảnh từ kỹ thuật mắt chim sử dụng WPM và IPM [4]*

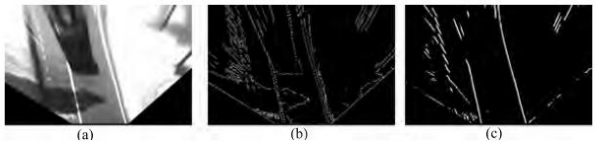
2.2.2. Phát hiện biên bằng kỹ thuật mắt chim

Bằng việc so sánh điểm ảnh hiện tại với điểm ảnh kế bên để phát hiện ra cạnh của của một hình ảnh.

$$\begin{aligned} B_{+m}(x,y) &= b(x,y) - b(x+m,y) \\ B_{-m}(x,y) &= b(x,y) - b(x-m,y) \end{aligned} \quad (2)$$

Theo công thức (3), nếu hiệu của 2 điểm ảnh này lớn hơn một mức ngưỡng thì có thể phát hiện ra cạnh của làn đường.

$$r(x,y) = \begin{cases} 1, & if\ B_{+m} \geq 0, B_{-m} \geq 0, and\ B_{+m} + B_{-m} \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

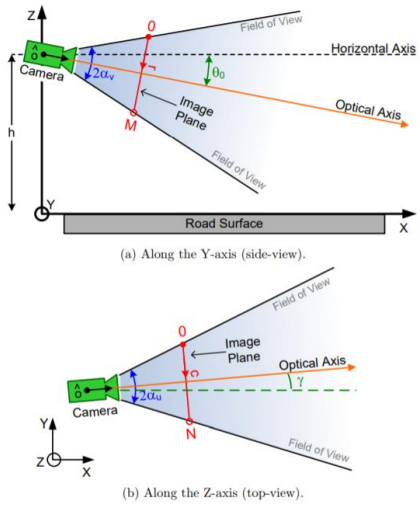


Hình 4. *Phát hiện biên [4]*

2.2.3. Dự đoán khoảng cách từ camera tới hình ảnh mắt chim

Nhắc lại một số thông số khi chuyển đổi hình ảnh từ tầm nhìn sang hình ảnh mắt chim.

- h là chiều cao của máy ảnh từ mặt đất,
- θ0: góc được hình thành bởi trục η và trục ngang
- Khẩu độ góc máy ảnh là 2αu theo hàng và 2αv theo cột
- γ0: góc được hình thành bởi phép chiếu của trục quang trên mặt phẳng xz
- n x m là độ phân giải của máy ảnh
- F là tiêu cự

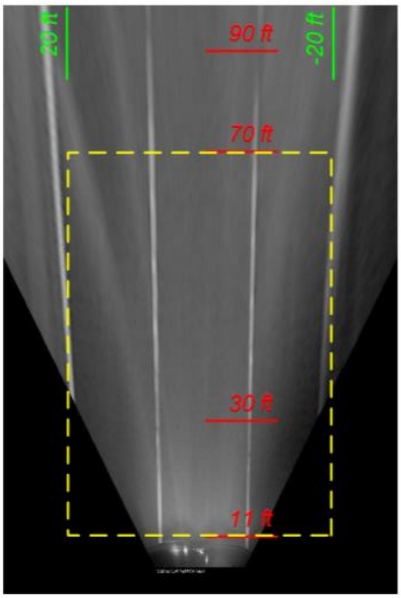


Hình 5. *Cấu hình camera*

Ta có:

$$\theta_0 = \tan^{-1}\left(\left[1 - 2\left(\frac{r_H - 1}{m-1}\right)\right]\tan\alpha_v\right) \quad (4)$$

Từ công thức (4) có thể ước tính được vị trí của vật thể đang di chuyển trên đường với camera đơn gắn cố định với góc nghiêng xác định trước, được kí hiệu là rH.



Hình 6. *Vị trí (rH) của vật thể theo chiều ngang trong hình ảnh camera*

Các vị trí của các khoảng cách khác nhau phía trước xe trong hình ảnh mắt chim. Khu vực được sử dụng để xử lý tiếp theo được thể hiện bằng hình chữ nhật màu vàng chấm đứt nét.

2.3. Đặc trưng kênh tổng hợp (Aggregate Channel Features – ACF)

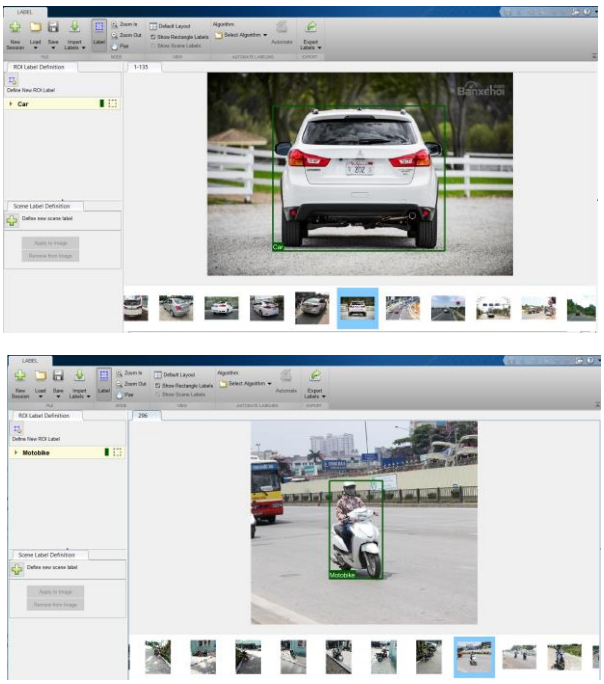
Đặc trưng kênh tổng hợp (Aggregate Channel Features – ACF) được đề xuất bởi Dollar [3] sử dụng việc tra cứu điểm ảnh trong các kênh tổng hợp để giảm thời gian trích đặc trưng mà không cần xây dựng các ảnh tích phân. Hiệu quả phối hợp của các kênh đặc trưng và việc chiến lược lựa chọn đặc trưng bằng cách xếp chồng các bộ phân loại làm các khung cửa sổ ACF hoạt động tốt hơn hầu hết các phương pháp khác về cả độ chính xác, tốc độ xử lý và có cách thực hiện đơn giản hơn.

2.3.1. Đặc điểm

a. Tập dữ liệu huấn luyện

Các dữ liệu huấn luyện (training data) được tạo từ Image Label Toolbox do Matlab hỗ trợ. Trong bài báo này, tập dữ liệu bao gồm:

- Tập dữ liệu huấn luyện của xe máy: 207 bức ảnh trong đó có một số bức ảnh có nhiều hơn một chiếc xe máy.
- Tập dữ liệu huấn luyện của xe hơi: 331 bức ảnh trong trong đó có một số bức ảnh có nhiều hơn một chiếc xe hơi.

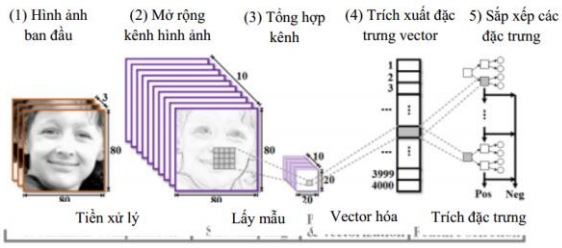


Hình 7. *Tracking đối tượng cho dữ liệu huấn luyện xe máy và xe hơi*

b. Mô hình thực hiện

Với hình I ngõ vào được tính toán với với một vài kênh C = Ω(I), tổng hợp mỗi khối pixel trong C, làm phẳng các kênh có độ phân giải thấp.

Các đặc trưng được tra cứu trong các pixel đơn lẻ trong các kênh tổng hợp. Tăng cường được sử dụng để huấn luyện và kết hợp trên các tính năng này (pixel) theo dạng cây để phân biệt đối tượng từ nền và sử dụng phương pháp tiếp cận đa cửa sổ trượt. Với sự lựa chọn thích hợp của các kênh và cấu trúc thiết kế đáng tin cậy, ACF đạt được hiệu quả tốt trong phát hiện người (hình 8).



Hình 8. *Mô hình ACF trong nhận dạng khuôn mặt người [3]*

Kênh (Channels)

ACF sử dụng 10 kênh như sau:

- Biên độ dốc (gradient magnitude): 1 kênh, độ lớn: $M(x,y)=\sqrt{I_x^2+I_y^2}$. Với I_x và

I_y là giá trị cường độ màu theo trục x và y tại vị trí thứ (x, y).

- Hướng dốc (gradient histogram): 6 kênh, từ 0 ~ 30; 31 ~ 60; 61 ~ 90; 91 ~ 120; 121 ~ 150; 151 ~ 180 độ).
- Kênh màu LUV (3 kênh): L – độ sáng, U và V – giá trị của màu trong ảnh. Kênh màu LUV ít bị ảnh hưởng từ các điều kiện sáng khác nhau.

Trước khi tính toán 10 kênh, hình ảnh ngõ vào I được làm phẳng với một bộ lọc có kích thước [1 2 1]/4.

2.4. Thuật toám RANSAC

Từ tập dữ liệu ban đầu, ta sẽ có hai loại dữ liệu nhiễu và không nhiễu (outlier và inlier), vì thế ta phải đi tính toán để tìm ra mô hình tốt nhất cho tập dữ liệu. Việc tính toán và chọn ra mô hình tốt nhất sẽ được lặp đi lặp lại k lần, với giá trị được chọn sao cho đủ lớn để đảm báo xác suất p (thường rơi vào giá trị 0.99) của tập dữ liệu mẫu ngẫu nhiên không chứa dữ liệu nhiễu.

Nếu gọi u là ước lượng dữ liệu không nhiễu thì v = 1 – u là ước lượng dữ liệu nhiễu và m là số lượng dữ liệu đầu vào cần xây dựng mô hình. Khi đó ta có:

$$1-p=\left(1-u^m\right)^k \tag{5}$$

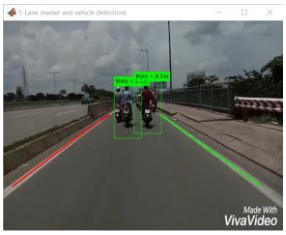
k sẽ được tính theo công thức:

$$k=\frac{\log(1-p)}{\log\left(1-(1-v)^m\right)} \tag{6}$$

Kết quả thu được sẽ là mô hình cần xây dựng phù hợp nhất với dữ liệu đầu vào, tập các dữ liệu nhiễu và tập các dữ liệu không nhiễu.

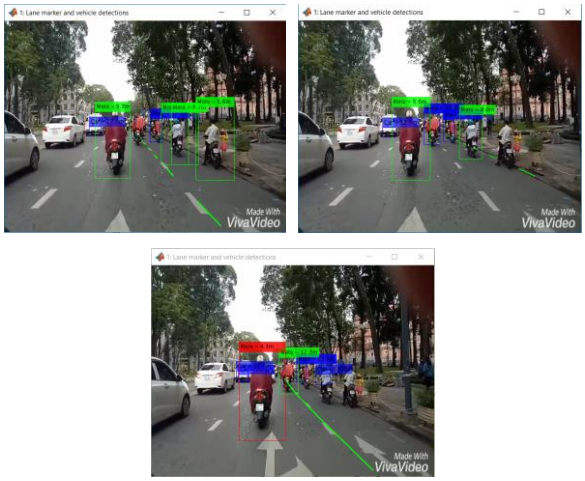
3. KẾT QUẢ

Video 1: Trên đường Phạm Văn Đồng, được quay lúc 12h trưa, với vị trí đặt của camera khoảng 1m7 tính từ mặt đất, góc nghiêng khoảng 10 độ.



Video 2: trên đường Khu Công Nghệ cao quận 9, được quay lúc 16h, với vị trí đặt của camera khoảng 2m tính từ mặt đất, góc nghiêng khoảng (12) độ.

Video 3: trên đường Lê Duẩn, trong điều kiện hơi âm u (có mưa), được quay lúc 16h, với vị trí đặt của camera khoảng 2m tính từ mặt đất, góc nghiêng khoảng (4) độ.



Video 4: trên đường Nam Kỳ Khởi Nghĩa, trong điều kiện hơi âm u (có mưa), được quay lúc 16h, với vị trí đặt của camera khoảng 2m tính từ mặt đất, góc nghiêng khoảng (4) độ.

Trường hợp nhận dạng chưa chính xác: làn đường có thể bị phát hiện chưa đúng do phụ thuộc vào cảm biến máy ảnh trong việc tạo hình ảnh mắt chim; nhận dạng sai giữa phương tiện là xe hơi hay xe máy hay không do dữ liệu huấn luyện không nhiều.

Đánh giá độ chính xác

Để đánh giá độ chính xác của kết quả nhận dạng, tác giả sử dụng thông số ROC (Reciver Operating Characteristic0, để xác định là phần nhận được có tín hiệu hay chỉ là do nhiễu.

- *True possitive (TP)*: Nếu kết quả từ dự đoán là có xe máy và giá trị thực tế cũng là có xe máy – Phân loại chính xác
- *True negatives (TN)*: Xảy ra khi cả kết quả dự đoán và giá trị thực tế là không có xe máy.
- *False positives (FP):* Nếu giá trị thực tế là không có xe máy thì nó được cho có xe máy – phân loại không chính xác
- *False negatives (FN):* Khi kết quả dự đoán là không có xe máy trong khi giá trị thực tế là có xe máy.

Đánh giá độ chính xác theo công thức (7):

$$accuracy = \frac{TP + TN}{P + N} \quad (7)$$

Bảng 1. *Thống kê độ chính xác của 4 video clip trong vấn đề nhận dạng xe hơi, xe máy và làn đường (khảo sát khoảng trên 20 frames)*

	TP	TN	FP	FN
Video 1	15	20	0	5
	ACC = 87.5 %			
Video 2	18	10	10	2
	ACC = 70 %			
Video 3	15	13	7	5
	ACC = 70 %			
Video 4	16	16	4	4
	ACC = 80 %			

4. KẾT LUẬN

4.1. Kết luận

Đề tài ban đầu đã áp dụng được kỹ thuật ACF trong nhận dạng một vài đối tượng tham gia giao thông đó là xe máy và xe hơi trên một số đoạn đường ở Việt Nam nhưng vẫn chưa áp dụng cho biển báo và các phương tiện hay các vật thể khác. Đề tài đã sử dụng kỹ thuật mắt chim đảo tầm nhìn của hình ảnh – IPM để phát hiện làn đường. Đề tài cũng dự đoán được khoảng cách từ camera đơn đến đối tượng phía trước.

Đề tài đã sử dụng Camera Calibration Toolbox trong việc hiệu chỉnh camera và Image Label Toolbox để tạo dữ liệu huấn luyện. Ngoài ra, đề tài cũng sử dụng thuật toán RANSAC trong việc loại bỏ nhiễu và sử dụng thông số ROC để đánh giá độ chính xác.

4.2. Hạn chế và hướng phát triển

Đề tài chỉ mới áp dụng cho những video đã được ghi lại bằng camera, chưa thực hiện trên camera thực tế khi tham giao thông (thực hiện online). Đề tài chưa thực hiện vào ban đêm hay đường đi dưới bóng cây (trong điều kiện thiếu ánh sáng). Đề tài sử dụng camera điện thoại nên phần cảm biến máy ảnh thay đổi liên tục theo điều kiện chiếu sáng, cũng như camera chưa chống rung và góc quay thay đổi liên tục nên gây ra có nhiều trường hợp nhận dạng chưa chính xác. Dữ liện huấn luyện còn hạn chế cũng làm độ chính xác chưa cao. Phần dự đoán khoảng cách chỉ sử dụng camera đơn nên cần sử dụng stereo camera để có kết quả chính xác. Dựa vào những hạn chế này, đề tài có thể định hướng được hướng cải thiện và phát triển cho những nghiên cứu tiếp theo áp dụng trong lĩnh vực lái xe tự động, giao thông thông minh.

TÀI LIỆU THAM KHẢO

[1] Mathwork, "*Visual Perception Using Monocular Camera*", Matlab link: https://www.mathworks.com/help/driving/examples/visual-perception-using-monocular-came

ra.html, 2017.

[2] Bin Yang, Junjie Yan, Zhen Lei, Stan Z. Li, "*Aggregate Channel Features for Multi-view Face Detection*", Center for Biometrics and Security Research & National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, China, IEEE, pp. 3-4, 2014.

[3] Piotr Dollar, Ron Appel, Serge Belongie, and Pietro Perona. "*Fast Feature Pyramids for Object Detection*", IEEE, pp. 9-10, 2014.

[4] Ruyi Jiang, Mutsuhiro Terauchi, Reinhard Klette, Shigang Wang, and Tobi Vaudrey, "*Low-level Image Processing for Lane Detection and Tracking*", Shanghai Jiao Tong University, Shanghai, China, Hiroshima International University, Japan, The University of Auckland, Auckland, New Zealand, pp. 3-7, 2009.

[5] Hsu-Yung Cheng, Bor-Shenn Jeng, Pei-Ting Tseng, and Kuo-Chin Fan, "*Lane Detection With Moving Vehicles in the Traffic Scenes*", IEEE, 2006.

[6] G. Loy, Nick Barnes, "*Fast shape-based road sign detection for a driver assistance system*", IEEE/RSJ International Confrence on Intelligent Robots and System, pp. 70-75, 2004.

[7] Mathworks, "*Single Camera Calibrator App*", Matlab Link: https://www.mathworks.com/help/vision/ug/single-camera-calibrator-app.html, 2015.

Tác giả chịu trách nhiệm bài viết:

Họ tên: Nguyễn Thụy Tô Hoài Nghi

Đơn vị: Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh

Điện thoại: 0356002243

Email: hoainghi.dvt09@gmail.com

BÀI BÁO KHOA HỌC

THỰC HIỆN CÔNG BỐ THEO QUY CHẾ ĐÀO TẠO THẠC SỸ

Bài báo khoa học của học viên

có xác nhận và đề xuất cho đăng của Giảng viên hướng dẫn

Bản tiếng Việt ©, TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH và TÁC GIẢ

Bản quyền tác phẩm đã được bảo hộ bởi Luật xuất bản và Luật Sở hữu trí tuệ Việt Nam. Nghiêm cấm mọi hình thức xuất bản, sao chụp, phát tán nội dung khi chưa có sự đồng ý của tác giả và Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh.

ĐỂ CÓ BÀI BÁO KHOA HỌC TỐT, CẦN CHUNG TAY BẢO VỆ TÁC QUYỀN!

Thực hiện theo MTCL & KHTHMTCL Năm học 2018-2019 của Thư viện Trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh.