

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN TOÁN ỨNG DỤNG & TIN HỌC

—o0o—



ỨNG DỤNG NHẬN DIỆN VÀ ĐẾM SỐ CÂY CỌ
SỬ DỤNG PHƯƠNG PHÁP HAAR-BASED CASCADE CLASSIFIER

ĐỒ ÁN 3

Chuyên ngành: Toán Tin

Giáo viên hướng dẫn	:	LÊ KIM THƯ
Sinh viên thực hiện	:	NGÔ TRƯỜNG GIANG
Lớp	:	TOÁN TIN K57
SHSV	:	20121584

Hà Nội - 06/2017

NHẬN XÉT CỦA THẦY HƯỚNG DẪN

1. Mục đích và nội dung của đề án

2. Kết quả đạt được

3. Ý thức làm việc của sinh viên ...

Hà Nội, ngày 05 tháng 05 năm 2017

Giảng viên hướng dẫn
(Ký và ghi rõ họ tên)

Mục lục

1	Lời cảm ơn	5
2	Lời nói đầu	6
3	Giới thiệu và đặt vấn đề	8
3.1	Viễn thám, ảnh vệ tinh độ phân giải cao, ảnh UAV	8
3.2	Máy học cơ bản và ứng dụng trong thị giác máy tính	10
3.3	Ứng dụng xử lý ảnh viễn thám vào nhận diện và đếm số cây cọ dầu	12
4	Phương pháp thực hiện	14
4.1	Phương pháp nhận diện đối tượng Haar Cascade	14
4.1.1	Đặc trưng dạng Haar và ảnh nguyên (Integral Image)	14
4.1.2	Nhận xét	15
4.1.3	Phương pháp xây dựng máy học phân lớp dựa trên AdaBoost	16
4.1.4	Phương pháp xây dựng máy học phân lớp hiệu quả dạng Cascade	17
5	Kết quả thực hiện	22
5.1	Vùng nghiên cứu và dữ liệu	22
5.2	Sử dụng thư viện OpenCV để luyện máy học nhận diện và đếm cây cọ	22
5.3	Mô tả chương trình	26
5.4	Kết quả và ước lượng độ chính xác	26
6	Kết luận	28
	Chỉ mục	29

Danh sách hình vẽ

1	Tình hình thị trường dầu cọ trong 5 năm trở lại đây	6
2	Kĩ thuật cửa sổ trượt	13
3	Một số đặc trưng dạng Haar cơ bản	15
4	Tổng cường độ sáng trong hình chữ nhật	16
5	Mô tả các kết quả của phân lớp nhị phân	20
6	Mô tả của một máy học nhận diện dạng cascade	21
7	ảnh vùng trồng cọ thứ nhất	23
8	ảnh vùng trồng cọ thứ hai	23
9	Tập ảnh mẫu dương	24
10	Tập ảnh mẫu âm	24
11	Hiển thị các đặc trưng dạng Haar được lựa chọn của một giai đoạn trong máy học	25
12	Giao diện ứng dụng tự động đếm số cây cọ trong ảnh	26
13	Ảnh các vùng dữ liệu được sử dụng để kiểm định độ chính xác của chương trình	27

Danh sách bảng

1	Thuật toán AdaBoost cho học phân lớp	18
2	độ chính xác của các phương pháp trên các dữ liệu kiểm thử	28

1 Lời cảm ơn

Lời đầu tiên, em xin chân thành cảm ơn các thầy giáo trong Trường Đại học Bách Khoa Hà Nội, cùng các thầy cô trong Viện Toán ứng dụng và Tin học, đã dành tâm huyết truyền đạt những kiến thức quý báu cho chúng em trong suốt những năm tháng học em tại trường.

Với lòng biết ơn sâu sắc, em xin cảm ơn cô Lê Kim Thư đã giúp đỡ em rất nhiều trong quá trình thực hiện đề án này.

Em cũng xin cảm ơn gia đình và bạn bè đã động viên, giúp đỡ em rất nhiều trong thời gian em làm đề án.

Cuối cùng em xin chúc các thầy cô giáo trong Trường Đại Học Bách Khoa Hà Nội lời chúc sức khỏe và thành đạt.

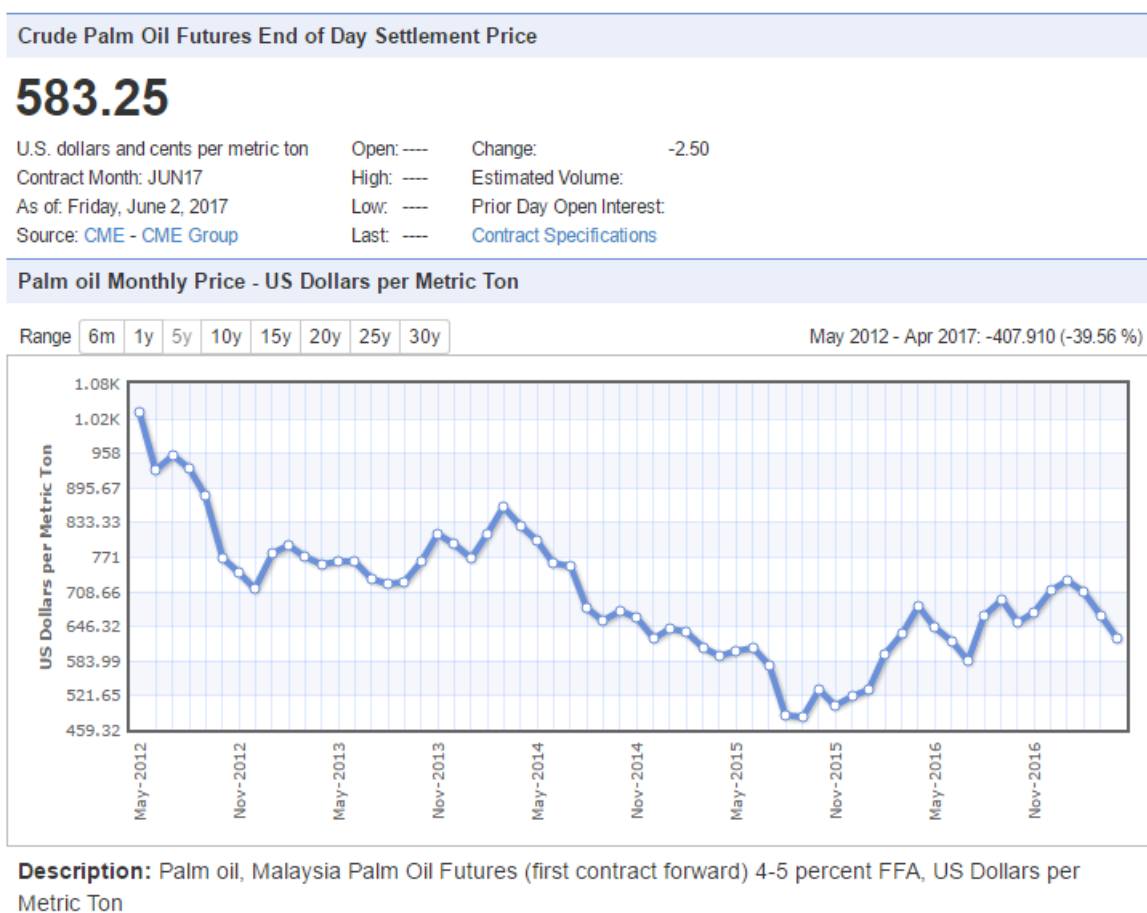
Hà Nội, ngày 05 tháng 05 năm 2017

Ngô Trường Giang

2 Lời nói đầu

Cây cọ dầu là loại cây công nghiệp có vai trò quan trọng trong kinh tế của Malaysia, đất nước đứng thứ hai thế giới về sản lượng dầu cọ, với sản lượng năm 2014 đạt 19.5 tỉ tấn. Nguồn lợi kinh tế đem lại cho đất nước này từ ngành công nghiệp dầu cọ ước tính 16.1 tỉ USD vào năm 2015, chiếm 5-6% GDP của cả đất nước.

Diện tích và sản lượng trồng cọ được kiểm soát bởi các hộ trồng cọ nhỏ lẻ (chiếm khoảng 40%) và các công ty dầu cọ lớn (chiếm khoảng 60%) (nguồn cleanmalaysia [7]).



Hình 1: Tình hình thị trường dầu cọ trong 5 năm trở lại đây (nguồn index-mundi[5])

Việc quản lý và kiểm soát được ngành trồng cây cọ dầu ở Malaysia mang

một ý nghĩa quan trọng, giúp dự báo được giá dầu cọ trên thị trường, tránh tình trạng cung vượt quá cầu khiến giá dầu cọ trên thị trường giảm mạnh như trong những năm gần đây. Giám sát được quá trình phát triển, và ước tính được sản lượng trồng cọ của mỗi vùng là những bước đầu tiên trong quản lý trồng

Số cây cọ trong một vùng trồng cọ chính là một thông tin cơ bản giúp phát hiện ra những vùng trồng cọ mới, theo dõi được sự phát triển của cây cọ và ước tính được sản lượng dầu cọ, chính là bước đầu trong quản lý nông nghiệp trồng cọ.

Với sự phát triển của công nghệ viễn thám (remote sensing), dữ liệu ảnh vệ tinh độ phân giải cao về mặt đất đang càng ngày càng phổ biến. Thực tế đã có nhiều bài báo nghiên cứu giải quyết bài toán đếm số cây cọ trong ảnh vệ tinh hoặc ảnh máy bay không người lái; nhiều phương pháp khác nhau được sử dụng như phương pháp học sâu sử dụng mạng neuron convolution [2], phương pháp Extreme learning machine trên nguồn ảnh chụp từ máy bay không người lái [3].

Trong báo cáo này, tôi sẽ trình bày phương pháp sử dụng ảnh viễn thám của vùng Jalan Bukit Kuantan, Malaysia, xây dựng chương trình tự động nhận diện và đếm số lượng cây cọ có trong ảnh chụp một vùng trồng cọ. Phương pháp được tôi sử dụng là phương pháp nhận diện đối tượng Haar-based Cascades Classifier, đề xuất bởi P. Viola và M. Jones[1] năm 2001.

Tôi đã sử dụng phương pháp này trên tập dữ liệu ảnh cây cọ dầu trích ra được từ ảnh viễn thám, để đào tạo một máy học phân lớp có khả năng nhận diện và đếm số cây cọ với độ chính xác tương đối tốt. Tôi đã sử dụng một số chương trình trong thư viện mã nguồn mở OpenCV [6] để đào tạo máy học này.

3 Giới thiệu và đặt vấn đề

3.1 Viễn thám, ảnh vệ tinh độ phân giải cao, ảnh UAV

Theo Wikipedia)[8], viễn thám là ngành khoa học về quan sát và thu thập thông tin của những đối tượng và hiện tượng diễn ra mà không cần liên hệ trực tiếp với đối tượng đó, do đó trái ngược với quan sát thực địa.

Viễn thám được sử dụng trong rất nhiều lĩnh vực khác nhau, gồm có địa lý, những ngành trong khoa học trái đất (sinh thái học, thủy văn học, hải dương học, địa chất học, v.v...), trong quân sự, kinh tế, và nông nghiệp, và nhiều ngành khác nữa.

Gần đây, khái niệm viễn thám thường được sử dụng để nhắc đến việc sử dụng các vệ tinh hoặc máy bay cùng với các công nghệ cảm biến hiện đại nhằm nhận diện và phân loại các vật thể trên trái đất.

Viễn thám có thể chia ra hai loại chính là viễn thám chủ động (tín hiệu được phát ra từ vệ tinh hoặc máy bay và bị phản xạ ngược trở lại hoặc bị hấp thụ bởi các đối tượng, kết quả thu lại được bằng các cảm biến) và viễn thám bị động (thu nhận sự phản chiếu ánh sáng mặt trời bởi các vật thể vào các cảm biến).

Từ những cảm biến này, qua quá trình số hóa và các thuật toán xử lý, các nhà khoa học viễn thám thu được thông tin dưới dạng ảnh viễn thám. Ảnh viễn thám có nhiều đặc tính kỹ thuật rất khác so với các ảnh chụp bởi máy ảnh quang thông thường, khi dải bước sóng mà các cảm biến thu được thường rộng hơn rất nhiều so với khoảng mắt người có thể nhìn thấy được, từ ảnh hồng ngoại cho đến ảnh RADAR. Chính điều này là ưu điểm của ảnh viễn thám, khi thu thập được nhiều thông tin hơn về các đối tượng trên trái đất.

Một số loại/nguồn ảnh viễn thám được sử dụng phổ biến: Landsat, NOAA-AVHRR, MODIS,...

Chất lượng của ảnh vệ tinh được quyết định bởi các độ phân giải sau:

- Độ phân giải không gian, kích thước của một pixel được ghi lại trong ảnh vệ tinh ứng với bao nhiêu mét trên thực địa.

- Độ phân giải phổ, số bước sóng của những giải tần số khác nhau tạo ra những băng khác nhau được ghi lại trong ảnh viễn thám.
- Độ phân giải giá trị thu được, khi số hóa các giá trị thu được bằng cảm biến, thường nằm trong phạm vi từ 8 cho tới 16 bit trong ảnh viễn thám.
- Độ phân giải thời gian là tần suất bay qua và thu được thông tin về một khu vực nhất định của vệ tinh hoặc máy bay và được sử dụng trong phân tích chuỗi thời gian.

Trong những năm gần đây, công nghệ cảm biến và khoa học hàng không vũ trụ đã phát triển mạnh mẽ, trở thành một ngành công nghiệp, khi có những công ty tự tay tạo ra những tên lửa cỡ nhỏ như Rocket Labs hay vệ tinh cỡ nhỏ như Dove của Planet Labs. Công nghệ viễn thám không còn là độc quyền của chính phủ các nước. Cùng với điều này, ảnh vệ tinh độ phân giải cao (very high resolution - VHR), với độ phân giải không gian nhỏ hơn 1m, đã trở nên phổ biến, được thu thập phủ khắp bề mặt trái đất, trong đó nổi bật là những vệ tinh mới và những loại ảnh mới như GeoEye-1 [9] và các vệ tinh WorldView-1, WorldView-2, WorldView-3 và WorldView-4 của DigitalGlobe [10].

Những ảnh vệ tinh độ phân giải cao đã đóng góp lớn trong công nghệ viễn thám hiện đại, khi mang đến ảnh chất lượng tốt hơn, các độ phân giải cao hơn, đa dạng thông tin hơn về trái đất, giúp người sử dụng ảnh đưa ra quyết định chính xác, cũng như góp phần xây dựng bản đồ thế giới đầy đủ hơn.

Bên cạnh ảnh vệ tinh độ phân giải cao, ảnh của thiết bị bay không người lái (unmanned aerial vehicle - UAV) cũng đang được sử dụng bởi cộng đồng viễn thám như một hệ thống thu thập dữ liệu thông tin địa lý mới không kém phần hiệu quả.

Tính linh động cao của những thiết bị UAV cho phép sự can thiệp ngay lập tức và đo lường có tương tác theo những yêu cầu cụ thể của khách hàng.

Trong thực tế, UAV có khả năng:

1. Cho phép theo dõi một vùng diện tích nhỏ với độ phân giải không gian

cực nhỏ (lên đến $1\text{cm}/\text{pixel}$).

2. Có thể di chuyển đến điểm mong muốn theo giới chỉ trong vài phút.
3. Cho phép thu thập thông tin theo chiều thời gian của cùng một vùng tại những khoảng thời gian xác định.

Trong một thời gian dài trước đây, ảnh UAV chỉ được sử dụng cho mục đích quân sự, nhưng với sự tiến bộ của công nghệ và giá thành giảm, những thiết bị bay không người lái trở thành một giải pháp thực tiễn cho nhiều ứng dụng khác nhau như giám sát nông nghiệp, phát hiện thay đổi bất thường, nông nghiệp chính xác hay đếm xe ô tô.

3.2 Máy học cơ bản và ứng dụng trong thị giác máy tính

Theo Wikipedia, **Máy học** (machine learning) là một lĩnh vực của trí tuệ nhân tạo (artificial intelligence) liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể.

Mỗi năm có hàng trăm thuật toán máy học mới ra đời, nhưng mọi thuật toán máy học đều có ba thành phần sau:

- **Mô hình biểu diễn tri thức.** Ví dụ có thể kể đến như cây quyết định, tập các luật, mạng neuron, support vector machines, v.v...
- **Phương pháp kiểm định chương trình,** nhằm kiểm tra xem liệu thuật toán có hoạt động tốt, đo lường hiệu quả của thuật toán.
- **Phương pháp tối ưu,** là phương pháp mà máy học tiếp cận đến mục tiêu của nó, ví dụ như phương pháp hướng giảm trong mô hình máy học dựa trên mạng neuron.

Về cơ bản, máy học có 4 loại sau:

- Học có giám sát: Dữ liệu được sử dụng để đào tạo máy học bao gồm có cả đầu ra mong muốn.
- Học không giám sát: Dữ liệu được sử dụng để đào tạo không bao gồm đầu ra mong muốn, ví dụ như phân cụm dữ liệu.
- Học nửa giám sát: Dữ liệu được sử dụng để đào tạo có một phần là đi kèm đầu ra mong muốn, một phần không.
- học tăng cường: máy học sẽ nhận được phần thưởng hoặc phạt sau một loạt các hành động tác động lên môi trường.

Cụ thể hơn, một số bài toán được nghiên cứu nhiều nhất trong lĩnh vực máy học bao gồm:

- Bài toán phân lớp, gán nhãn cho những giá trị đầu vào.
- Bài toán phân cụm, phân chia dữ liệu đầu vào thành các cụm hay các miền dựa trên thuộc tính của chúng.
- Bài toán hồi quy, tìm ra hàm biểu diễn quan hệ giữa dữ liệu đầu vào và dữ liệu đầu ra.

Một trong những lĩnh vực thường xuyên được áp dụng các kỹ thuật máy học nhiều nhất chính là thị giác máy tính (computer vision). Thị giác máy tính là một lĩnh vực liên ngành, quan tâm đến việc giúp máy tính đạt được khả năng hiểu được ảnh và video kỹ thuật số, từ đó trích xuất những thông tin hữu ích. Từ quan điểm kỹ thuật, thị giác máy tính tìm kiếm sự tự động hóa những tác vụ mà hệ thống thị giác của mắt người có thể làm được.

Một số tác vụ thông thường của hệ thống thị giác máy tính: nhận diện, phân tích chuyển động, khôi phục ảnh và tái tạo khung cảnh.

Nhận diện đối tượng là một hướng chính và được tập trung rất nhiều trong thị giác máy tính, đồng thời những kỹ thuật máy học được tạo ra để phục vụ cho tác vụ này cũng rất phong phú. Trong nhận diện đối tượng thì nhiệm vụ cơ bản là xác định được vị trí của một loại đối tượng cần tìm ở trong

ảnh. Nếu có nhiều đối tượng thuộc loại đó xuất hiện thì phải tìm được vị trí của tất cả các đối tượng này.

Bài toán đếm số cây cọ trong ảnh viễn thám chính là một dạng cụ thể của bài toán nhận diện đối tượng, ở đây là nhận diện và đếm số cây cọ xuất hiện.

Trong báo cáo này, tôi sẽ đưa về bài toán nhận diện và đếm số cây cọ trong ảnh thành bài toán xây dựng **máy học phân lớp nhị phân** (binary classification) (chỉ phân làm hai lớp có và không hay dương và âm), có khả năng xác định một ảnh có phải là ảnh chụp cây cọ hay không, hay còn gọi là khả năng nhận diện cây cọ trong ảnh.

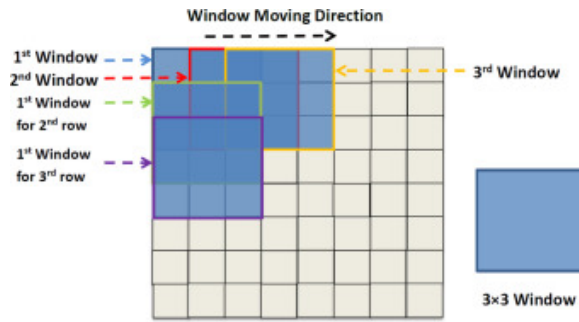
Sử dụng máy học phân lớp nhị phân này, kết hợp với phương pháp duyệt từng cửa sổ trong ảnh, slicing window, ta có một phương pháp tìm kiếm và đếm số cây cọ trong ảnh.

Trong phương pháp cửa sổ trượt (slicing window), mỗi cửa sổ là một khung hình chữ nhật với kích thước cho trước, nhỏ hơn kích thước của ảnh, sẽ được cho trượt đi theo chiều dọc và chiều ngang, mỗi lần một số lượng pixel cho trước, cho đến khi duyệt qua khắp các vị trí trong ảnh.

Những ảnh con thu được bằng việc cắt ảnh gốc sử dụng cửa sổ này sẽ được đưa vào máy học phân lớp nhị phân. Máy học nhận diện đối tượng lúc này sẽ kiểm tra xem liệu có cây cọ ở trong ảnh con này hay không. Nếu có thì số lượng cây cọ đếm được sẽ tăng lên một đơn vị, ngược lại thì dịch chuyển cửa sổ sang vị trí tiếp theo. Khi một cửa sổ đã duyệt hết bức ảnh, ta có thể thay đổi kích thước cửa sổ và thực hiện duyệt lại lần nữa để tìm kiếm những đối tượng có kích thước nhỏ hơn hoặc lớn hơn.

3.3 Ứng dụng xử lý ảnh viễn thám vào nhận diện và đếm số cây cọ dầu

Cọ dầu là cây công nghiệp phát triển trong vùng khí hậu nóng ẩm. Dầu cọ được chiết xuất từ thịt của quả cây cọ dầu trở thành nguyên liệu thô cho rất nhiều sản phẩm mà đứng đầu là dầu ăn vì sản lượng của dầu cọ rất lớn so với



Hình 2: Kỹ thuật cửa sổ trượt trong nhận diện đối tượng.

các loại cây cho dầu khác như hướng dương, đậu nành hay cải dầu.

Dầu cọ có nhu cầu rất lớn trên thị trường khiến cho sự phát triển của diện tích trồng cọ đã tăng trưởng nhanh chóng trong vài thập kỉ gần đây.

Trong báo cáo này, tôi sẽ trình bày phương pháp đếm số cây cọ trong ảnh viễn thám độ phân giải cao, chụp một vùng trồng cọ ở thành phố Kuantan, Malaysia.

Ảnh vệ tinh độ phân giải cao sau khi được tiền xử lý, sẽ được dùng để lấy mẫu những ảnh con chứa cây cọ dầu và những ảnh con không chứa cây cọ dầu. Những ảnh mẫu được tạo ra từ quá trình tiền xử lý và lấy mẫu sẽ được dùng để luyện máy học giúp tự động nhận diện và đếm số cây cọ dầu có ở trong ảnh.

Với phương pháp nhận diện đối tượng Haar-based Cascade Classifier được đề xuất bởi P. Viola và M. Jones [1], máy học thu được bằng cách áp dụng phương pháp này trong có khả năng đếm nhanh chóng số lượng cây cọ dầu trong ảnh.

Cuối cùng, tôi viết một ứng dụng desktop, sử dụng thư viện OpenCV[6] và ngôn ngữ lập trình Python [13], thực thi phương pháp nhận diện đối tượng Haar-based Cascade Classifier, để đếm số cây cọ có trong một bức ảnh viễn thám khác của vùng khác thuộc Kuantan, Malaysia.

4 Phương pháp thực hiện

Sau đây, tôi sẽ trình bày phương pháp học máy được sử dụng trong bài này để nhận diện và đếm số cây cọ dầu trong ảnh viễn thám, phương pháp Haar-Based Cascade Classifier do P. Viola và M. Jones[1] đề xuất.

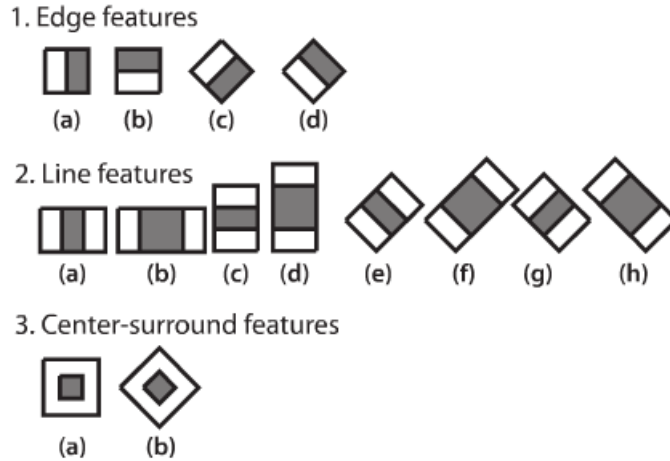
4.1 Phương pháp nhận diện đối tượng Haar Cascade

4.1.1 Đặc trưng dạng Haar và ảnh nguyên (Integral Image)

Thủ tục nhận diện đối tượng ở trong một bức ảnh phụ thuộc vào giá trị thu được của những đặc trưng thu được từ ảnh đó. Có rất nhiều lý do cho việc sử dụng đặc trưng thay vì từng pixel ảnh. Lý do chung nhất là những đặc trưng này có khả năng mã hóa những tri thức học được từ tập dữ liệu luyện. Lý do thứ hai là hệ thống phân lớp dựa trên đặc trưng thì nhanh hơn dựa trên từng pixel. Những đặc trưng dạng Haar là những đặc trưng của ảnh kỹ thuật số, có cái tên bắt nguồn từ sự tương tự với Haar wavelet, và là một trong những đặc trưng được sử dụng phổ biến trong nhận dạng đối tượng.

Papageorgiou et al. [1] đã đề cập đến việc sử dụng những đặc trưng dựa trên Haar wavelets thay vì sử dụng cường độ sáng tại mỗi điểm ảnh. Viola và Jones đã sử dụng ý tưởng này và phát triển những đặc trưng dạng Haar (Haar-like features). Một đặc trưng dạng Haar tính toán trên những vùng hình chữ nhật kề nhau tại một vị trí xác định trên cửa sổ nhận diện, tính tổng cường độ sáng (giá trị tại mỗi pixel trong ảnh grayscale) trong mỗi vùng và tính hiệu của những tổng này. Ví dụ, ta có một cơ sở dữ liệu hình ảnh về mặt người. Dễ dàng quan sát thấy trong tất cả các ảnh rằng vùng mắt thường tối hơn so với vùng má. Do đó một đặc trưng dạng Haar cho mặt người có thể là hai hình chữ nhật kề nhau nằm trên vùng mắt và vùng má. Vị trí của hai hình chữ nhật này được định nghĩa tương đối với cửa sổ nhận diện.

Để tính toán tổng cường độ sáng trong một hình chữ nhật nhanh chóng, Viola và Jones[1] đã sử dụng một dạng biểu diễn trung gian cho ảnh được gọi là ảnh nguyên (integral image). Ảnh nguyên có cùng kích thước với ảnh thực,



Hình 3: Một số đặc trưng dạng Haar cơ bản, (1) Các đặc trưng 2 hình chữ nhật, (2) Các đặc trưng 3 hình chữ nhật, (3) Các đặc trưng dạng trung tâm.

giá trị của ảnh nguyên tại mỗi điểm ảnh (x, y) bao gồm tổng cường độ sáng của số pixel nằm về phía bên trái và phía trên của (x, y) . Công thức tính giá trị ảnh nguyên:

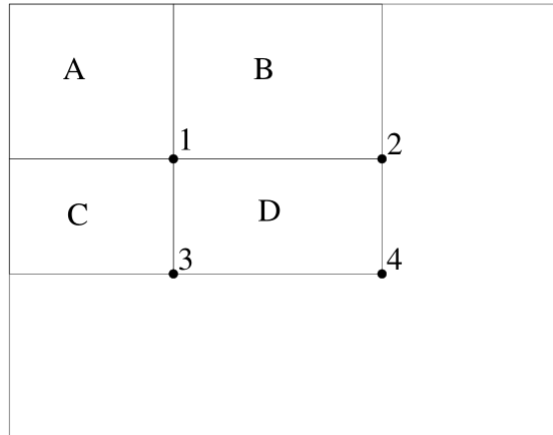
$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

Với $ii(x, y)$ là ảnh nguyên và $i(x, y)$ là ảnh gốc. Sử dụng ảnh nguyên, bất kì hình chữ nhật có cũng có thể tính tổng cường độ sáng nhanh bằng 4 giá trị của ảnh nguyên tại 4 đỉnh của hình chữ nhật, thời gian tính toán là hằng số. Đây cũng chính là ưu điểm của đặc trưng dạng Haar so với các đặc trưng khác, thời gian tính toán rất nhanh.

4.1.2 Nhận xét

Những đặc trưng dạng Haar chỉ giới hạn trong các hình chữ nhật, và chỉ có 3 hướng cơ bản: dọc, ngang và chéo 45° , nên những đặc trưng này thường rất yếu trong nhận diện ảnh. Tuy nhiên, tập hợp của những đặc trưng này, lại cung cấp một sự thể hiện tốt của đối tượng cần nhận diện.

Với độ phân giải của cửa sổ nhận diện là 24×24 , số lượng đặc trưng dạng



Hình 4: Tổng cường độ sáng trong hình chữ nhật D được tính bằng $ii(4) + ii(1) - ii(2) - ii(3)$

Haar cũng đã khá lớn, lên đến hơn 180, 000 đặc trưng. Đa phần trong số chúng là không có ý nghĩa. Mặc dù các đặc trưng có thể tính toán rất nhanh, tính toán trên tập tất cả đặc trưng này vẫn rất tốn thời gian. Giả thiết được Viola và Jones[1] đặt ra bằng thực nghiệm là chỉ có một số lượng rất ít trong số những đặc trưng này có thể kết hợp để tạo ra một máy học phân lớp hiệu quả. Do đó, ta cần lựa chọn ra những đặc trưng hiệu quả trong nhận diện đối tượng và xây dựng một máy học phân lớp dựa trên tập những đặc trưng hiệu quả này.

4.1.3 Phương pháp xây dựng máy học phân lớp dựa trên AdaBoost

Cho một tập những đặc trưng và một tập dữ liệu luyện được gán nhãn dương (có đối tượng cần nhận diện) và âm (không có đối tượng cần nhận diện), bất kỳ phương pháp máy học phân lớp nào cũng có thể dùng để học hàm phân lớp hai loại dữ liệu này. Viola và Jones[1] đã lựa chọn một phương pháp tương tự như AdaBoost để lựa chọn ra một tập nhỏ những đặc trưng và luyện một máy học phân lớp.

Thuật toán ở đây được thiết kế để lựa chọn ra những đặc trưng tốt nhất cho việc phân lớp những ảnh dương và âm. Cho mỗi đặc trưng, một hàm

ngưỡng phân lớp tối ưu được lựa chọn để tối thiểu hóa số ảnh bị phân nhầm lớp. Hàm phân lớp này kí hiệu là $h_j(x)$ của đặc trưng f_j , với ngưỡng θ_j và parity p_j để chỉ chiều của bất phương trình, có dạng:

$$h_j(x) = \begin{cases} 1 & \text{nếu } p_j f_j(x) < p_j \theta_j, \\ 0 & \text{ngược lại.} \end{cases}$$

Ở đây x là một cửa sổ kích thước 24×24 của ảnh.

Sơ đồ của quá trình boosting được trình bày trong bảng 1.

4.1.4 Phương pháp xây dựng máy học phân lớp hiệu quả dạng Cascade

Phần này sẽ mô tả một thuật toán xây dựng một máy học phân lớp dạng thang cascade hay dạng tầng, để tăng hiệu năng của quá trình nhận diện đồng thời giảm thời gian tính toán.

Trước hết ta nhắc đến một số khái niệm trong phân lớp nhị phân được sử dụng ở phần này.

Trong phân lớp nhị phân (binary classification), khi phân lớp một tập dữ liệu, hàm phân lớp được sử dụng cùng với một ngưỡng (threshold) được xác định để phân lớp. Hai lớp trong phân lớp nhị phân thường được gọi là lớp âm (negative) và lớp dương (positive). Trong nhận diện đối tượng, lớp dương thường là tập những đối tượng cần nhận diện, còn lớp âm thường là tập những đối tượng không phải đối tượng cần nhận diện.

Giả sử gọi hàm phân lớp nhị phân là hàm $f(x)$ và ngưỡng của phân lớp là t . khi $f(x) > t$, dữ liệu đầu vào x được phân vào lớp dương, ngược lại x được phân vào lớp âm.

Tùy vào cách lựa chọn ngưỡng t mà ta thu được kết quả phân lớp với các đo lường về độ chính xác khác nhau. Một số khái niệm được định nghĩa dựa trên phân lớp nhị phân:

- Một dữ liệu được gọi là True Positive (TP) nếu nó là dữ liệu dương và được nhận diện là thuộc lớp dương.

Bảng 1: Thuật toán AdaBoost cho học phân lớp. Mỗi vòng lặp lựa chọn ra một đặc trưng từ danh sách những đặc trưng.

- Cho một tập ảnh mẫu $(x_1, y_1), \dots, (x_n, y_n)$ với $y_i = 0, 1$ ứng với mẫu là âm (không chứa đối tượng cần nhận diện) hay dương (có chứa đối tượng cần nhận diện).
- Khởi tạo các trọng số $w_{1,i} = \frac{1}{2m}$ với $y_i = 0$ và $w_{1,i} = \frac{1}{2l}$ với $y_i = 1$, trong đó m và l lần lượt là số mẫu dương và mẫu âm.
- For $t = 1, \dots, T$:

1. Chuẩn hóa những trọng số:

$$w_{t,i} \rightarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

sao cho vectơ trọng số w_t là phân phối xác suất.

2. với mỗi đặc trưng, j , luyện một hàm phân lớp h_j . Tỷ lệ lỗi được tính ứng với w_t ,

$$\epsilon_j = \sum_i w_j |h_j(x_i) - y_i|$$

3. Chọn một hàm phân lớp h_t với tỷ lệ lỗi thấp nhất ϵ_t

4. cập nhật các trọng số:

$$w_{t+1,i} = w_{t,i} \beta_t^{(1-e_i)}$$

trong đó $e_i = 0$ nếu x_i được phân lớp đúng, $e_i = 1$ nếu ngược lại, và $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- Hàm phân lớp cuối cùng được tính dựa trên các h_t được lựa chọn ra:

$$h(x) = \begin{cases} 1 & \text{nếu } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t, \\ 0 & \text{ngược lại.} \end{cases}$$

với $\alpha_t = \log \frac{1}{\beta_t}$

- Một dữ liệu được gọi là True Negative (TN) nếu nó là dữ liệu âm và được nhận diện là thuộc lớp âm.
- Một dữ liệu được gọi là False Positive (FP) nếu nó là dữ liệu âm nhưng được nhận diện thuộc lớp dương.
- Một dữ liệu được gọi là False Negative (FN) nếu nó là dữ liệu dương nhưng được nhận diện thuộc lớp âm.

Các tỉ lệ True Positive, True Negative, False Positive và False Negative được tính như sau:

Tỉ lệ true positive (true positive rate):

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

Tỉ lệ true negative (true negative rate):

$$TNR = \frac{TN}{TN + FP} \quad (2)$$

Tỉ lệ false negative (false negative rate):

$$FNR = \frac{FN}{FN + TP} \quad (3)$$

Tỉ lệ false positive (false positive rate):

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

Ý tưởng của phân lớp dạng cascade là sử dụng những máy học phân lớp đơn giản được xây dựng để loại bỏ phần lớn những cửa sổ âm (tỉ lệ false negative gần tới 0), trước khi sử dụng những máy học phân lớp phức tạp để đạt được tỉ lệ false positive thấp (loại bỏ những cửa sổ âm bị nhận diện là đối tượng).

Ngưỡng của một máy học phân lớp có thể điều chỉnh để đạt được tỉ lệ false negative gần tới 0, để đảm bảo không có cửa sổ nào có đối tượng bị bỏ sót, hoặc ngược lại, loại bỏ đi đa phần những cửa sổ âm.

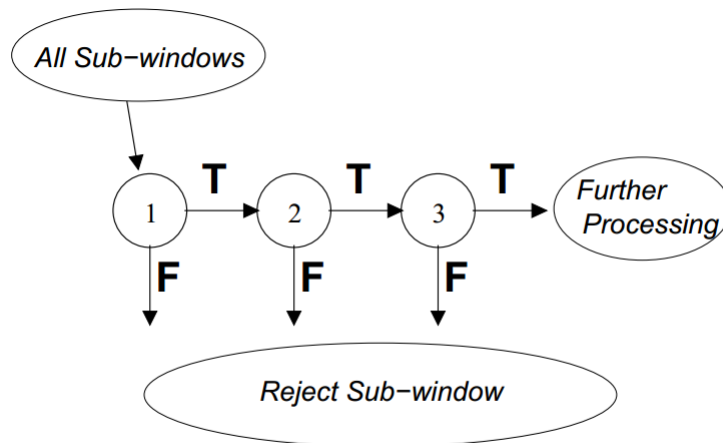
		Actual	
		True	False
Classifier	True	True Positive	False Positive
	False	False Negative	True Negative

Hình 5: Mô tả các kết quả của phân lớp nhị phân.

Dạng tổng quát của quá trình nhận dạng là một dạng đơn giản của cây quyết định, mà ta gọi là cascade, gồm có các giai đoạn (stage). Mỗi một giai đoạn ứng với một máy học phân lớp nhận diện cây cọ, được tạo ra bằng thuật toán dạng AdaBoost trình bày trong bảng 1. Những kết quả dương từ máy học phân lớp đầu tiên sẽ được đưa vào máy học phân lớp thứ hai và tiếp tục phân lớp. Kết quả dương của máy học phân lớp thứ hai sẽ được đưa vào máy học thứ ba, cứ thế tiếp diễn. Những cửa sổ bị các máy học phân lớp nhận diện là âm sẽ ngay lập tức bị loại bỏ khỏi lần phân lớp tiếp theo.

Cấu trúc của máy học nhận dạng cascade phản ánh thực tế rằng trong một bức ảnh, đa phần cửa sổ là âm (không phải là đối tượng cần nhận diện), vì thế nó cố gắng loại bỏ những cửa sổ âm này ở những giai đoạn sớm nhất có thể, những cửa sổ dương sẽ được phân lớp dương lần lượt qua các mọi máy học phân lớp của cascade.

Trong hầu hết các trường hợp, nhiều đặc trưng được sử dụng hơn sẽ giúp đạt được tỉ lệ nhận diện cao hơn và tỉ lệ false positive thấp hơn. Nhưng đồng thời máy học phân lớp có nhiều đặc trưng sẽ tốn nhiều thời gian để tính toán



Hình 6: Mô tả của một máy học nhận diện dạng cascade. Một chuỗi những máy học phân lớp được áp dụng cho những cửa sổ. Máy học phân lớp đầu tiên sẽ loại bỏ một số lượng lớn những ví dụ âm với xử lý rất đơn giản (chỉ có một vài đặc trưng được sử dụng). Những máy học phân lớp tiếp theo sẽ loại bỏ những cửa sổ âm khác nhưng phức tạp hơn (nhiều đặc trưng hơn). Sau một vào bước xử lý, số cửa sổ phải xử lý giảm đi đáng kể. Phần cuối là xử lý về sau, có thể sử dụng một hệ thống nhận diện khác hoặc tiếp tục thêm các máy học phân lớp vào.

hơn. Khi luyện một máy học phân lớp dạng cascade, ta quan tâm đến: 1. số máy học phân lớp, 2. số đặc trưng trong mỗi máy học, 3. ngưỡng của mỗi máy học, được điều chỉnh để tối thiểu hóa số đặc trưng cần tính toán nhưng vẫn đảm bảo độ chính xác yêu cầu.

Trong thực tế, một framework rất đơn giản được sử dụng để tạo ra một máy học phân lớp có hiệu năng cao. Mỗi một giai đoạn của máy học cascade được luyện bằng cách thêm vào những đặc trưng vào máy học giai đoạn này cho đến khi tỉ lệ nhận diện vào tỉ lệ false positive đạt được yêu cầu. Các giai đoạn được thêm vào cho tới khi tỉ lệ false positive và tỉ lệ nhận diện tổng thể đạt yêu cầu.

5 Kết quả thực hiện

5.1 Vùng nghiên cứu và dữ liệu

Dữ liệu được sử dụng trong báo cáo này là hai ảnh được chụp bởi máy bay không người lái (UAV), chụp một vùng trồng cọ thuộc vùng Jalan Bukit Kuantan, Malaysia tại hai địa điểm khác nhau.

5.2 Sử dụng thư viện OpenCV để luyện máy học nhận diện và đếm cây cọ

Trong bộ thư viện mã nguồn ở OpenCV đã có sẵn chương trình để luyện máy học phân lớp dạng cascade, được sử dụng để nhận dạng mặt người như một ví dụ áp dụng lý thuyết mà Viola Jones đã đề xuất.

Tôi đã sử dụng lại chương trình này, chạy với bộ dữ liệu cây cọ dầu thu thập được, thu được một máy học có khả năng nhận diện và đếm số cây cọ dầu có trong ảnh.

Sau đây là các bước của quá trình luyện máy học nhận diện cây cọ dầu:

1. Chuẩn bị dữ liệu. Tại bước này, tôi sử dụng công cụ mã nguồn mở QGIS [11] để thao tác trên dữ liệu thông tin địa lý, trích xuất ra ảnh những cây cọ từ ảnh của cả vùng trồng cọ.
 - Bằng mắt, tôi chấm các điểm trên giao diện của QGIS, mỗi điểm này ứng với tâm của một cây cọ, đồng thời mỗi điểm được lưu dưới dạng tọa độ địa lý của điểm đó trên bản đồ.
 - Tập những điểm này được lưu lại dưới dạng file Shapefile [?], một định dạng lưu dữ liệu thông tin địa lý dạng vectơ, cụ thể ở đây là các điểm ứng với các tọa độ địa lý.
 - Bằng ngôn ngữ lập trình Python [13] và thư viện xử lý dữ liệu địa lý vectơ và ảnh viễn thám GDAL [14], tôi viết mã xử lý file shapefile đã tạo ra, trích lấy tọa độ các điểm trong file này, chuyển



Hình 7: ảnh vùng trồng cọ thứ nhất, được sử dụng để làm dữ liệu luyện máy học nhận dạng cây cọ



Hình 8: ảnh vùng trồng cọ thứ hai, được sử dụng để làm dữ liệu kiểm tra độ chính xác của máy học nhận dạng

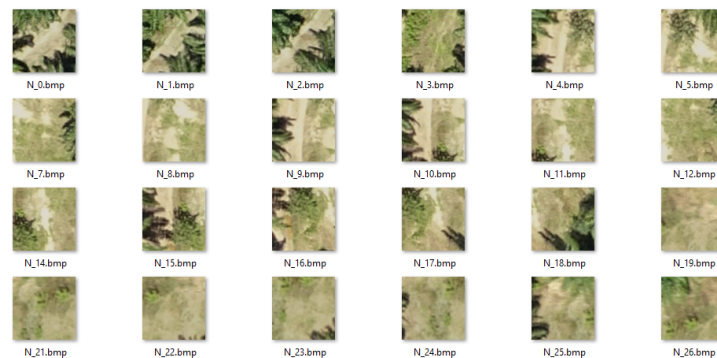
các điểm này sang hệ tọa độ của bức ảnh viễn thám, lựa chọn lấy một khung hình kích thước 80x80 lấy điểm này làm tâm, cắt ảnh

nằm trong khung hình đó ra, lưu lại dưới dạng file ảnh. Thực hiện thao tác này với từng điểm trong file shapefile để thu được danh sách những ảnh chứa cây cọ để làm mẫu dương.

Lặp lại các bước trên, nhưng lựa chọn các vùng không phải là cây cọ, ta thu được tập dữ liệu ảnh mẫu âm.



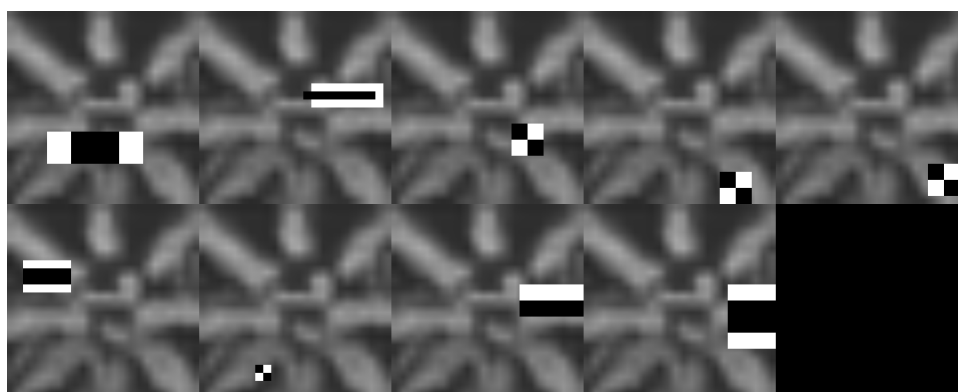
Hình 9: Tập ảnh mẫu dương



Hình 10: Tập ảnh mẫu âm

2. Cấu hình tham số và chạy máy học Cascade Classifier trong OpenCV. Thư viện OpenCV đã có bộ công cụ giúp luyện một máy học nhận diện đối tượng theo phương pháp của Viola và Jones. Chi tiết về cách sử dụng công cụ này có trong link sau: [docs.opencv.org/3.2.0/dc/d88-/tutorial_traincascade.html]. Bộ công cụ này gồm có 3 chương trình chính, chạy trong cửa sổ dòng lệnh: **opencv_createsamples**, **opencv_traincascade** và **opencv_visualisation**.

- Chương trình **opencv_createsamples** được sử dụng để đọc dữ liệu mẫu dương, lựa chọn các khung hình có chứa đối tượng cần nhận diện và lưu lại dưới định dạng .vec để sử dụng bởi chương trình **opencv_traincascade**.
 - Chương trình **opencv_traincascade** nhận đầu vào là danh sách các dữ liệu mẫu âm, file định dạng .vec được tạo bởi chương trình **opencv_createsamples**, và các tham số khác của thuật toán, xuất ra một file mô tả mô hình cascade của máy học nhận dạng, định dạng xml.
 - File kết quả xml được sử dụng bởi hàm của thư viện OpenCV để tạo ra máy học nhận diện, dựa trên các mô tả và tham số luyện được. Máy học này sẽ nhận đầu vào là ảnh chứa đối tượng cần nhận diện, đầu ra là các khung hình chữ nhật bao quanh đối tượng cần nhận diện.
3. Hiển thị kết quả thu được từ máy học. Chương trình **opencv_visualization** cho phép hiển thị kết quả của quá trình luyện máy học nhận diện cây cọ, là những đặc trưng được lựa chọn, và thứ tự của những đặc trưng này trong máy học phân lớp lên trên một ảnh mẫu cây cọ được chọn.

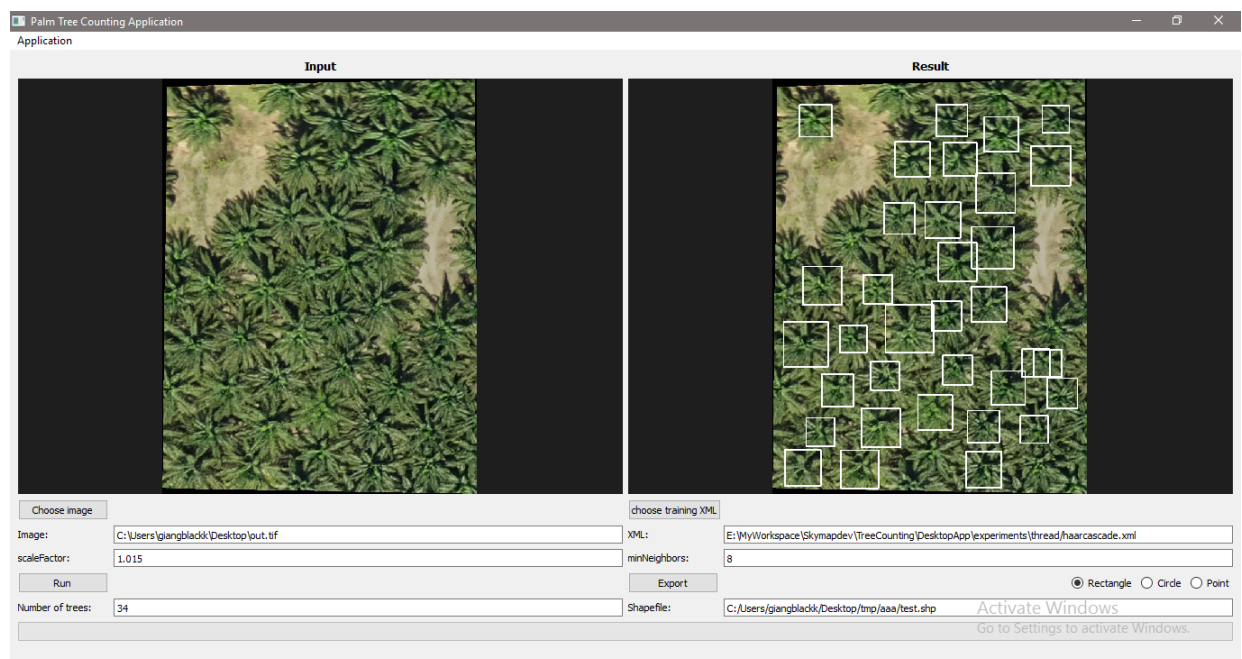


Hình 11: Hiển thị các đặc trưng dạng Haar được lựa chọn của một giai đoạn (giai đoạn 11) trong máy học đào tạo được.

5.3 Mô tả chương trình

Sau khi đã có kết quả của quá trình luyện máy học nhận diện là file xml mô tả máy học và các đặc trưng được lựa chọn, tôi tiếp tục sử dụng thư viện OpenCV, ngôn ngữ lập trình Python và thư viện làm giao diện PySide để xây dựng chương trình desktop tự động nhận diện và đếm số cây cọ trong một bức ảnh.

Chương trình cho phép lựa chọn ảnh để đếm cây với nút **Choose Image**, Sau khi nhấn nút **Run**, chương trình sẽ tự động xử lý và hiển thị kết quả ở khung hình bên phải. Các cây được nhận diện sẽ được khoanh bằng các hình vuông màu trắng. Nút **Export** cho phép xuất kết quả nhận diện ra dưới định dạng file shapefile.



Hình 12: Giao diện ứng dụng tự động đếm số cây cọ trong ảnh.

5.4 Kết quả và ước lượng độ chính xác

Độ chính xác của quá trình nhận dạng được đánh giá định lượng thông qua 3 chỉ số: precision, recall và overall accuracy, được tính dựa trên kết quả

nhận diện được khi so sánh với kết quả cây thực sự có trong ảnh.

Công thức của các chỉ số này tham khảo từ [2] như sau:

$$\text{Precision} = \frac{\text{Số đối tượng nhận diện được đúng là cây cọ}}{\text{Tổng số đối tượng nhận diện}} \quad (5)$$

$$\text{Recall} = \frac{\text{Số đối tượng nhận diện được đúng là cây cọ}}{\text{Số cây cọ thực sự có trong ảnh}} \quad (6)$$

$$\text{Overall Accuracy} = \frac{\text{Precision} + \text{Recall}}{2} \quad (7)$$

Một nhận diện được cho là nhận diện đúng cây cọ khi tâm của nhận diện đó cách tâm của một cây cọ thực sự không quá 10 pixel.

Bảng 2 cung cấp kết quả về các chỉ số đánh giá độ chính xác định lượng của chương trình với các dữ liệu mẫu được sử dụng để kiểm tra. Ta nhận thấy tỉ lệ chính xác này là tương đối tốt.



Hình 13: Ảnh các vùng dữ liệu được sử dụng để kiểm định độ chính xác của chương trình.

Bảng 2: độ chính xác của các phương pháp trên các dữ liệu kiểm thử

Dữ liệu	Precision(%)	Recall(%)	Overall Accuracy(%)
Vùng 1	87.5	89.09	88.30
Vùng 2	96.15	89.28	92.7
Vùng 3	94.87	82.22	88.545
Vùng 4	90	81.81	85.9
Vùng 5	77.78	82.35	80.06
Vùng 6	93.33	77.78	85.56
Vùng 7	94.53	98.25	96.39
Vùng 8	96.96	98.76	195.72

6 Kết luận

Trong báo cáo này, tôi đã trình bày quá trình sử dụng phương pháp nhận diện đối tượng Haar-based Cascade Classifier đề xuất bởi P. Viola và M. Jones, nhằm xây dựng máy học nhận diện và đếm số cây cọ trong ảnh viễn thám độ phân giải cao thu thập được từ vùng Jalan Bukit Kuantan, Malaysia. Trong quá trình này tôi có sử dụng đến thư viện OpenCV để luyện máy học nhận diện cây cọ và ngôn ngữ lập trình Python, cùng các thư viện mã nguồn mở hỗ trợ xử lý dữ liệu thông tin địa lý khác. Bằng trực quan có thể thấy kết quả của máy học nhận diện vẫn còn chưa đủ tốt, khi có những vùng nhỏ trong ảnh có cây cọ nhưng lại không hề nhận diện được.

Trong những nghiên cứu tiếp theo, tôi đề xuất sử dụng một số hướng giúp cải thiện hiệu quả của phương pháp như: tăng kích thước bộ dữ liệu luyện máy học, gồm có tập dữ liệu có cây cọ và tập dữ liệu không có cây cọ; sử dụng ảnh từ những cây chưa nhận diện được, thêm vào tập dữ liệu luyện để luyện lại; sử dụng các phương pháp máy học khác như convolution neural network. Do năng lực và thời gian có hạn, báo cáo còn nhiều sai sót mong nhận được góp ý thêm, tôi xin chân thành cảm ơn.

Chỉ mục

ảnh của thiết bị bay không người
lái, 9
ảnh vệ tinh độ phân giải cao, 9
ảnh viễn thám, 8

Máy học, 10
Thị giác máy tính, 11
Viễn thám, 8

Tài liệu

- [1] C. Papageorgiou, M. Oren, T. Poggio, "A general framework for object detection", International Conference on Computer Vision, 1998.
- [2] Weijia Li, Haohuan Fu, Le Yu, Arthur Cracknell, "Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote", Remote Sensing, 2017, Vol. 9, ISSN: 2072-4292 Sensing Images
- [3] Salim Malek, Yakoub Bazi, Naif Alajlan, Haikel AlHichri, Farid Melgani, "Efficient Framework for Palm Tree Detection in UAV Images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 12, pp. 4692-4703, Dec. 2014.
- [4] P. Viola and M. Jones, *Rapid object detection using a boosted cascade of simple features*, in IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1. IEEE, 2001, pp. 1-511.
- [5] Palm oil Price, <http://www.indexmundi.com/commodities/?commodity=palm-oil>
- [6] OpenCV library, <http://opencv.org>
- [7] Just how Big is Malaysia's Palm Oil Industry?, <http://cleanmalaysia.com/2015/12/09/just-how-big-is-malaysias-palm-oil-industry/>
- [8] Remote Sensing, https://en.wikipedia.org/wiki/Remote_sensing
- [9] GeoEye-1 Satellite Image, <http://www.satimagingcorp.com/gallery/geoeye-1/>
- [10] Digital Globe, <https://www.digitalglobe.com/>
- [11] QGIS - A Free and Open Source Geographic Information System, <http://www.qgis.org/en/site/>

- [12] Shapefiles, <http://doc.arcgis.com/en/arcgis-online/reference/shapefiles.htm>
- [13] Python language, <https://www.python.org/>
- [14] GDAL - Geospatial Data Abstraction Library, <http://gdal.org/>