

## MÔN HỌC: XỬ LÝ ẢNH

# ĐỀ TÀI: NHẬN DIỆN PHƯƠNG TIỆN GIAO THÔNG TRÊN VIDEO

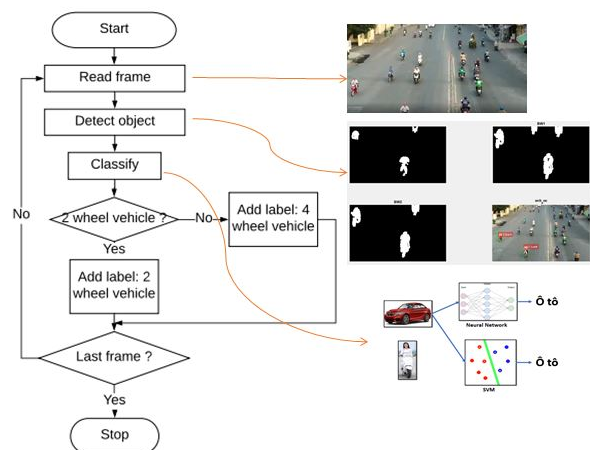
GVHD: ĐẶNG NGUYỄN CHÂU

Nguyễn Công Kỳ 1511676  
Lê Quang Huy 1511234  
Trần Đức Minh 1512022  
Mai Văn Thời 1513319

**Tóm tắt nội dung**—Trong báo cáo này, nhóm chúng em thực hiện đề tài "Nhận dạng phương tiện giao thông trên video". Mục tiêu là viết chương trình Matlab, đọc 1 ảnh phương tiện giao thông trên video và phân loại. Kết quả, nhóm đã thực hiện được các bước xử lý ảnh đơn giản để nhận dạng các đặc trưng.

## I. GIỚI THIỆU ĐỀ TÀI

Giao thông tại Việt Nam được đánh giá là khá phức tạp với mật độ phương tiện cao, thêm vào đó là đa dạng các loại hình phương tiện giao thông. Trong thời buổi hiện tại, việc xử lý, thu thập dữ liệu để đánh giá, xử lý vi phạm giao thông là thực sự cần thiết. Trong khuôn khổ đề tài này, chúng tôi đề xuất phương pháp nhận diện xe 2 bánh và xe 4 bánh trên video giao thông Việt Nam sử dụng phương pháp phân lớp Support Vector machine và Mạng Neural Network có áp dụng thêm phương pháp trích đặc trưng Histogram Of Gradient.



Hình 1: lưu đồ giải thuật của đề tài

## II. LƯU ĐỒ GIẢI THUẬT

### A. Tạo tập dữ liệu training (dataset):

#### 1) Tạo bộ dataset cho xe hai bánh:

Phía trước lấy 40 bức hình khác nhau của xe máy chụp ở phía trước  
Phía sau lấy 99 bức hình khác nhau của xe máy chụp ở phía sau  
Bên hông lấy 100 bức hình khác nhau của xe máy được chụp ngang:

#### 2) Tạo bộ dataset cho xe bốn bánh:

Phía trước lấy 50 bức hình khác nhau của xe ô tô chụp ở phía trước  
Phía sau lấy 78 bức hình khác nhau của xe ô tô chụp ở phía sau  
Bên hông lấy 116 bức hình khác nhau của xe ô tô được chụp ngang:

**Xe 2 bánh - Trước: 40**      **Xe 4 bánh - Trước: 50**  
**Xe 2 bánh - Sau: 99**      **Xe 4 bánh - Sau: 78**  
**Xe 2 bánh - Ngang: 100**      **Xe 4 bánh - Ngang: 116**



Hình 2: tập dữ liệu dataset tự tạo



Hình 3: tách 2 xe gần nhau

## B. Tiền xử lý cho các đối tượng chồng lên nhau

### C. bước 1:

Tách ngưỡng  $\rightarrow$  Dilation

Tách cạnh  $\rightarrow$  Dilation

D. bước 2: Dán nhãn và tách đối tượng ra khỏi ảnh, dùng HOG để trích đặc trưng cho ảnh

E. bước 3: Dùng mạng Neural Network và SVM training để phân loại các phương tiện giao thông

F. bước 4: gán nhãn cho từng loại phương tiện giao thông

## III. CÁC LÝ THUYẾT LIÊN QUAN

### A. Phương pháp trích đặc trưng HOG

HOG(histogram of oriented gradients) là một feature descriptor được sử dụng trong computer vision và xử lý hình ảnh, dùng để detect một đối tượng. Các khái niệm về HOG được nêu ra từ năm 1986 tuy nhiên cho đến năm 2005 HOG mới được sử dụng rộng rãi sau khi Navneet Dalal và Bill Triggs công bố những bổ sung về HOG. Hog tương tự như các biểu đồ edge orientation, scale-invariant feature transform descriptors(như sift, surf,...), shape contexts nhưnghog được tính toán trên một lưới dày đặc các cell và chuẩn hóa sự tương phản giữa các block để nâng cao độ chính xác. Hog được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một object trong ảnh Bài toán tính toán Hog thường gồm 5 bước:

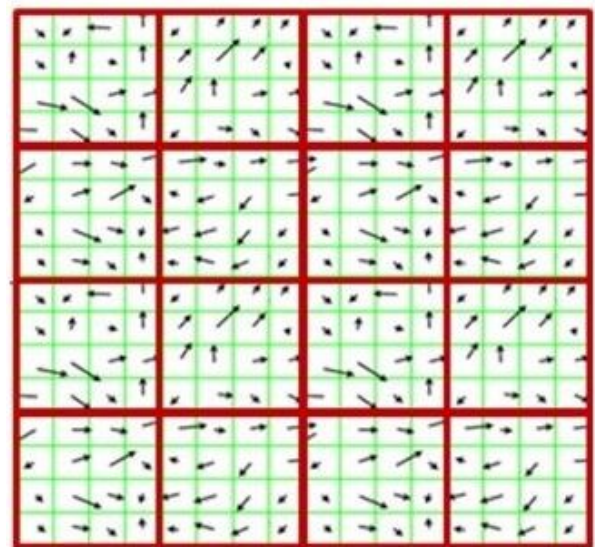
- Chuẩn hóa hình ảnh trước khi xử lý
- Tính toán gradient theo cả hướng x và y .
- Lấy phiếu bầu cùng trọng số trong các cell
- Chuẩn hóa các block
- Thu thập tất cả các biểu đồ cường độ gradient định hướng để tạo ra feature vector cuối cùng

#### 1) Chuẩn hóa hình ảnh trước khi xử lý:

Bước chuẩn hóa này hoàn toàn không bắt buộc, nhưng trong một số trường hợp, bước này có thể cải thiện hiệu suất của bộ mô tả HOG

#### 2) Tính toán gradient:

Để lấy được hình ảnh gradient, chúng ta sẽ sử dụng tích chập, sau khi có các ảnh gradient, chúng ta có thể tính toán cường độ gradient của hình ảnh, cuối cùng là định hướng của gradient cho mỗi pixel trong hình ảnh ban đầu

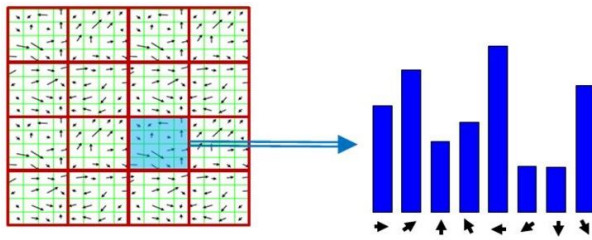


Hình 4: Tính toán gradients cho từng pixel

#### 3) Lấy votes trong mỗi cell:

Bây giờ chúng ta cần chia hình ảnh của chúng ta thành các cell và block Một cell là một vùng hình chữ nhật được xác định bởi số điểm ảnh thuộc mỗi cell. Ví dụ: nếu ta có một hình ảnh 128 x 128 với pixelpercell = 4 x 4 thì sẽ có 32 x 32 = 1024 cell, pixelpercell = 32 x 32, sẽ có 4 x 4 = 16 cell. Với mỗi cell trong bức ảnh, ta cần xây dựng 1 biểu đồ cường độ gradient. Mỗi pixel sẽ được vote vào vào biểu đồ, trọng số của mỗi vote chính là cường độ gradient tại pixel đó Cuối cùng, mỗi pixel đóng góp một phiếu bầu có trọng số vào biểu đồ - trọng lượng của phiếu chỉ đơn giản là cường độ gradient |G| tại pixel đó.Lúc này, chúng ta có thể thu thập và ghép các biểu đồ này để tạo ra feature vector

cuối cùng. Tuy nhiên, ta sẽ chuẩn hóa các block để có được kết quả tốt hơn



Hình 5: Tính toán histogram cho từng vùng của bức ảnh

#### 4) Chuẩn hóa các block:

Một lần nữa, ta cần chia các block giống như chia cell ở phía trên. Đơn vị của ta không còn là các điểm ảnh nữa mà là các cell. Người ta thường sử dụng hoặc 2 x 2 hoặc 3 x 3 cell per block có được độ chính xác hợp lý trong hầu hết các trường hợp. Các block này sẽ chồng lên nhau. Ví dụ: ta có 3x3 cells và cell per block = 2x2 thì ta sẽ có 4 block. Tiếp đến, ta sẽ tiến hành thu thập và ghép các histogram của cell trong block.

#### 5) Sử dụng trong bài toán object recognition:

HOG được implement trong opencv và scikit-image. Tuy nhiên, việc sử dụng HOG trong scikit-image linh hoạt hơn rất nhiều trong opencv. Bài toán mình đặt ra là, trong một bức ảnh có nhiều chữ số có kích thước khác nhau, làm thế nào để xác định được các chữ số đó. Đầu tiên, mình sử dụng opencv2 để load ảnh, resize và chuyển ảnh sang màu grayscale. Sau đó lấy ra các hình chữ nhật xung quanh các ký tự, ta sẽ tìm hog feature trên mỗi mảng điểm ảnh của ký tự được lấy ra. Cuối cùng ta xây dựng một model KNN.

### B. Support Vector Machine

SVM sử dụng để tìm ra một siêu phẳng (hyperplane) - chính là cái đường cong cong như hình trên đó. Nhưng thử tưởng tượng trong không gian nhiều chiều hơn chẳng hạn, nó có thể là một mặt cầu, mặt bầu dục... Tóm lại mục đích của cái siêu phẳng đó là phân tách tập dữ liệu thành hai phần riêng biệt - tư tưởng của bài toán phân lớp.

Trong không gian 2 chiều, khoảng cách từ điểm A(x<sub>0</sub>, y<sub>0</sub>) tới đường thẳng  $w_1x + w_2y + b = 0$  :

$$\frac{|w_1x_0 + w_2y_0 + b|}{\sqrt{w_1^2 + w_2^2}}$$

Trong không gian 3 chiều, khoảng cách từ điểm A(x<sub>0</sub>, y<sub>0</sub>) tới mặt phẳng  $w_1x + w_2y + w_3z + b = 0$  :

$$\frac{|w_1x_0 + w_2y_0 + w_3z_0 + b|}{\sqrt{w_1^2 + w_2^2 + w_3^2}}$$

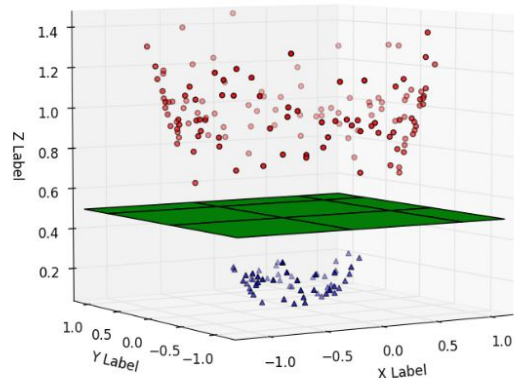
=> Khoảng cách từ một điểm X tới siêu phẳng có phương trình  $w_1x + w_2y + w_3z + \dots + b = 0$  :

$$\frac{|w_1x_0 + w_2y_0 + w_3z_0 + \dots + b|}{\sqrt{w_1^2 + w_2^2 + \dots + w_n^2}}$$

#### Margin trong SVM là gì?

Margin là khoảng cách giữa siêu phẳng đến 2 điểm dữ liệu gần nhất tương ứng với các phân lớp. Ví dụ quả táo quả lê đặt trên mặt bán, margin chính là khoảng cách giữa cây que và hai quả táo và lê gần nó nhất.

=> Điều quan trọng ở đây đó là phương pháp SVM luôn cố gắng cực đại hóa margin này, từ đó thu được một siêu phẳng tạo khoảng cách xa nhất. Nhờ vậy, SVM có thể giảm thiểu việc phân lớp sai (misclassification) đối với điểm dữ liệu mới đưa vào.



Hình 6: chuyển không gian từ 2 chiều sang 3 chiều

=> SVM là một phương pháp hiệu quả cho bài toán phân lớp dữ liệu. Nó là một công cụ đặc lực cho các bài toán về xử lý ảnh, phân loại văn bản, phân tích quan điểm. Một yếu tố làm nên hiệu quả của SVM đó là việc sử dụng Kernel function khiến cho các phương pháp chuyển không gian trở nên linh hoạt hơn..

### IV. DEMO, NHẬN XÉT VÀ ĐÁNH GIÁ

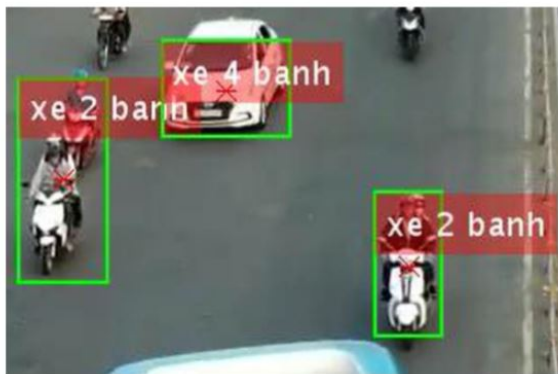
#### A. Đánh giá mạng đã Training và chạy demo

Sau khi chạy thử hai thuật toán HOG và SVM ta thu được bảng kết quả như sau:

1. Sử dụng SVM				2. Sử dụng Mạng Neural Network			
Đối tượng	Số mẫu train	Số mẫu test	Số mẫu nhận diện đúng	Đối tượng	Số mẫu train	Số mẫu test	Số mẫu nhận diện đúng
Xe 2 bánh	90	30	30	Xe 2 bánh	90	30	21
Xe 4 bánh	90	30	18	Xe 4 bánh	90	30	26

Accuracy: 80%                      Accuracy: 75%

Hình 7: kết quả sau khi chạy thuật toán SVM và mạng NN



Hình 8: chạy demo

## B. Nhận xét ưu nhược điểm và đánh giá

### 1) Trích đặc trưng dùng HOG

#### : Ưu điểm:

- Rất mạnh mẽ.
- Rất hữu ích để biểu diễn các đối tượng có cấu trúc không thay đổi đáng kể.
- Rất chính xác trong Object Classification.

#### Nhược điểm:

- Feature Vector có thể trở nên rất lớn nếu không chú ý các tham số như orientations, pixelspercell, cellsperblock và kích thước input image.
- Khá chậm.
- Không hoạt động tốt nếu cấu trúc của đối tượng cần được mô tả bị thay đổi lớn (ví dụ: đối tượng bị xoay).

### 2) Phân loại bằng Neural Network

#### : Ưu điểm:

- Phân loại cùng lúc được nhiều đối tượng.

#### Nhược điểm:

- Với số mẫu lớn thì tốn bộ nhớ, cần cấu hình máy tốt.

### 3) Phân loại bằng SVM

#### : Ưu điểm:

- *Xử lý trên không gian số chiều cao*: SVM là một công cụ tính toán hiệu quả trong không gian chiều cao, trong đó đặc biệt áp dụng cho các bài toán phân loại văn bản và phân tích quan điểm nơi chiều có thể cực kỳ lớn
- *Tiết kiệm bộ nhớ*: Do chỉ có một tập hợp con của các điểm được sử dụng trong quá trình huấn luyện và ra quyết định thực tế cho các điểm dữ liệu mới nên chỉ có những điểm cần thiết mới được lưu trữ trong bộ nhớ khi ra quyết định
- *Tính linh hoạt*: phân lớp thường là phi tuyến tính. Khả năng áp dụng Kernel mới cho phép linh động giữa các phương pháp tuyến tính và phi tuyến tính từ đó khiến cho hiệu suất phân loại lớn hơn

#### Nhược điểm:

- *Bài toán số chiều cao*: Trong trường hợp số lượng thuộc tính (p) của tập dữ liệu lớn hơn rất nhiều so với số lượng dữ liệu (n) thì SVM cho kết quả khá tồi
- *Chưa thể hiện rõ tính xác suất*: Việc phân lớp của SVM chỉ là việc cố gắng tách các đối tượng vào hai lớp được phân tách bởi siêu phẳng SVM. Điều này chưa giải thích được xác suất xuất hiện của một thành viên trong một nhóm là như thế nào. Tuy nhiên hiệu quả của việc phân lớp có thể được xác định dựa vào khái niệm margin từ điểm dữ liệu mới đến siêu phẳng phân lớp mà chúng ta đã bàn luận ở trên

### 4) Đánh giá tổng quan

#### : Ưu điểm:

- Chương trình cơ bản đã phát hiện đúng và nhận diện được đối tượng
- Tập dataset tự tạo, gần với thực tế hơn các tập dataset trên mạng, từ đó cải thiện kết quả huấn luyện và nhận dạng.
- Có so sánh tính khả thi của 2 phương pháp với nhau.

- Kết hợp ngõ ra của cả 2 phương pháp để cho kết quả tối ưu hơn.

**Nhược điểm:**

- Vẫn còn sai số.
- Còn chưa phân biệt tốt các đối tượng bị phân cắt hoặc chồng chéo lên nhau.

**V. HƯỚNG TỐI ƯU, PHÁT TRIỂN ĐỀ TÀI:**

- Tăng cường thuật toán để nhận diện các vật bị phân cắt hoặc chồng chéo lên nhau.
- phát triển thêm để nhận diện phương tiện với mật độ cao hơn.
- Áp dụng thuật toán tiên tiến hơn.