

**ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH**



BÁO CÁO THỰC TẬP TỐT NGHIỆP

ƯỚC LƯỢNG SỐ LƯỢNG VÀ ĐỘ HỖN LOẠN CỦA ĐÁM ĐÔNG TỪ VIDEO

GVHD: TS. LÊ THÀNH SÁCH

GVPB: ThS. TRẦN GIANG SƠN

---o0o---

SVTH: NGUYỄN HOÀNG MINH (51001955)

TP. HỒ CHÍ MINH, 06/2014

LỜI CAM ĐOAN



Bản báo cáo thực tập tốt nghiệp này có tham khảo các tài liệu, bài báo, trang web như được trình bày ở mục Tài liệu tham khảo và ở mỗi tham khảo đều được trích dẫn nguồn gốc rõ ràng. Tôi xin cam đoan rằng ngoài những trích dẫn từ những tham khảo trên, toàn bộ nội dung trong bài báo cáo này là do tôi tự soạn thảo ra từ những kết quả kết quả nghiên cứu của riêng tôi, không sao chép từ bất kì tài liệu nào khác.

Tôi sẽ hoàn toàn chịu trách nhiệm xử lý theo qui định nếu có bất cứ sai phạm nào so với lời cam kết

Sinh viên thực hiện đề tài Nguyễn Hoàng Minh

LỜI CẢM ƠN



Đầu tiên, em xin gửi lời cảm ơn chân thành đến quý Thầy, Cô của Khoa Khoa học và Kỹ thuật Máy tính Trường Đại học Bách Khoa Thành phố Hồ Chí Minh đã tận tình chỉ bảo, truyền đạt nền tảng kiến thức vững chắc, kỹ năng cuộc sống cho sinh viên của trường nói chung và bản thân em nói riêng. Đặc biệt, em muốn gửi lời cảm ơn đến Thầy Lê Thành Sách là giảng viên hướng dẫn đã hết lòng hướng dẫn, giúp đỡ em suốt quá trình thực hiện đề tài.

Nền tảng kiến thức vững chắc và những kinh nghiệm quý báu em tích lũy được trong quá trình học tập từ quý Thầy, Cô của trường sẽ là hành trang không thể thiếu cho công việc của em sau này.

Một lần nữa, em trân trọng cảm ơn các Thầy cô, cô chú Cán bộ, Nhân viên của Trường Đại học Bách Khoa Thành phố Hồ Chí Minh đã tạo điều kiện môi trường học tập tốt nhất cho em và sinh viên của trường, cùng lời chúc sức khỏe, thành công và hạnh phúc.

TÓM TẮT



Vấn đề về giám sát đám đông ở các khu vực đông người như ga tàu điện ngầm, trung tâm thương mại, quảng trường, sân vận động,... nhằm kiểm soát an ninh, hạn chế rủi ro để đưa ra các quyết định giải quyết kịp thời. Việc nắm bắt số lượng và độ hỗn loạn của đám đông mang ý nghĩa to lớn để giải quyết vấn đề nêu trên. Hơn thế nữa, bài toán trên có thể được áp dụng vào tình hình giao thông ở Việt Nam. Việc ước lượng số lượng phương tiện trong ảnh thu được từ camera một cách tự động trên các tuyến đường mang lại những giải pháp cho vấn đề giao thông hiệu quả nhất.

Có 2 cách để nắm bắt số lượng người trong ảnh, đó là bằng phương pháp trực tiếp và phương pháp gián tiếp. Hướng tiếp cận được thực hiện trong đề tài này là ước lượng số lượng người theo phương pháp gián tiếp.

Cốt lõi của cách thức thực hiện sự ước lượng sẽ trình bày trong đề tài này là khai thác ảnh đã được lọc với bộ lọc Gabor ở nhiều kênh khác nhau. Việc khai thác sử dụng ảnh sau khi lọc với bộ lọc Gabor là chìa khóa quyết định độ chính xác sự ước lượng của hệ thống. Một yếu tố quan trọng là tính tổng quát khi áp dụng phương pháp vào thực tiễn. Tính tổng quát thể hiện rằng khi train mô hình học máy với 1 lượng dataset đủ nhiều, kết quả ước lượng các ảnh test bất kỳ trả về từ mô hình học máy phải đạt độ chính xác chấp nhận được. Đương nhiên, một số điều kiện ràng buộc nhất định xuất hiện kèm theo.

Trên cơ sở công trình nghiên cứu của các tác giả đi trước, sau nhiều tháng nghiên cứu và thí nghiệm, tôi đề xuất một số phương pháp giúp cải thiện độ chính xác, tăng tính tổng quát của hệ thống ước lượng. Độ chính xác phương pháp đề xuất đạt 97% trong thí nghiệm.

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN.....	ii
TÓM TẮT.....	iii
MỤC LỤC	iv
DANH MỤC BẢNG BIỂU	vi
DANH MỤC HÌNH ẢNH.....	vii
1. TỔNG QUAN ĐỀ TÀI.....	1
1.1 Giới thiệu đề tài.....	1
1.2 Ứng dụng của đề tài	2
1.3 Hướng tiếp cận.....	2
2. KIẾN THỨC NỀN TẢNG	3
2.1 Cân bằng sáng Histogram	3
2.2 Convolution.....	4
2.3 Phân cụm K-means	6
2.4 Gradient.....	7
3. Phương pháp ước lượng.....	9
3.1 Tổng quan các phương pháp khác	9
3.1.1 Phương pháp sử dụng hình dạng của nhóm người.....	9
3.1.2 Phương pháp thống kê đặc trưng góc chuyển động	9
3.1.3 Phương pháp sử dụng đặc trưng SURF.....	9
3.1.4 Nhận xét	10
3.2 Phương pháp Gabor filter	10
3.2.1 Tổng quan phương pháp	10
3.2.2 Thiết kế bộ lọc Gabor.....	11
3.3 Phương pháp 1 - Chia lưới.....	13

3.4	Phương pháp 2 - Hoán vị	15
3.5	Phương pháp 3 - Bag of words	15
3.6	Phương pháp 4 - Histogram	18
3.7	Phương pháp 5 - Gradient	21
4.	KẾT QUẢ THÍ NGHIỆM	23
4.1	Dataset	23
4.2	Kết quả so sánh	24
5.	NHẬN XÉT ĐÁNH GIÁ	27
5.1	Ưu điểm của các phương pháp	27
5.2	Nhược điểm của các phương pháp	27
6.	KẾT LUẬN	29
6.1	Kết luận	29
6.2	Kiến nghị	29
	TÀI LIỆU THAM KHẢO	30

DANH MỤC BẢNG BIỂU

Bảng 1. Thông tin dữ liệu thí nghiệm	24
Bảng 2. Độ chính xác của các phương pháp trên Data1, Data2, Data3	24
Bảng 4. Độ chính xác của các phương pháp khi sử dụng ảnh gradient	25
Bảng 5. So sánh các phương pháp với phương pháp Gradient	26

DANH MỤC HÌNH ẢNH

Hình 1. Ảnh, histogram trước và sau khi cân bằng sáng.....	4
Hình 2. Các pixel góc phía trên bên trái của ảnh	5
Hình 3. Xây dựng pixel nằm ngoài ảnh.....	5
Hình 4. Vùng mặt nạ nhân tương ứng	6
Hình 5. Minh họa giải thuật K-means	7
Hình 6. Ảnh gradient	8
Hình 7. Kiến trúc hệ thống phương pháp Gabor filter	11
Hình 8. Hình ảnh 24 bộ lọc Gabor (chẵn)	12
Hình 9. Ảnh sau khi đi qua kênh Gabor thứ nhất (lẻ)	13
Hình 10. Trích đặc trưng theo mean và variance của các ảnh đầu ra.....	13
Hình 11. Phương pháp chia lưới.....	14
Hình 12. Thiết lập mặt nạ	14
Hình 13. Lưới 3x4	14
Hình 14. Áp dụng chia lưới	15
Hình 15. Phương pháp hoán vị.....	15
Hình 16. Lưới trên vùng quan sát - phương pháp Bag of words.....	16
Hình 17. Phân cụm - phương pháp Bag of words	16
Hình 18. Các điểm trong hệ trục mean, variance	17
Hình 19. Trích xuất đặc trưng trên mỗi kênh - phương pháp Bag of words	17
Hình 20. Phân chia bin - phương pháp Histogram.....	18
Hình 21. Xây dựng bin - phương pháp Histogram.....	19
Hình 22. Bin tương ứng với góc - phương pháp Histogram	19
Hình 23. Trích xuất đặc trưng trên mỗi kênh - phương pháp Histogram.....	20
Hình 24. Phương pháp Gradient.....	21
Hình 25. Ảnh lọc Gabor (trái) và ảnh gradient tương ứng (phải)	22
Hình 26. Hình ảnh trong tập dataset.....	23
Hình 27. Biểu đồ so sánh - Gabor filter, Data1	25
Hình 28. Biểu đồ so sánh - Chia lưới, Data1	25

1. TỔNG QUAN ĐỀ TÀI

1.1 Giới thiệu đề tài

Vấn đề về kiểm soát đám đông tự động bằng hệ thống camera an ninh không còn xa lạ đối với các nước phát triển. Nó đang được chú trọng đầu tư nghiên cứu kỹ lưỡng nhằm có thể giải quyết triệt để 1 số vấn đề liên quan khi tình trạng người trở nên đông đúc. Video đám đông được phân tích chuẩn xác sẽ mang lại những thông tin hữu ích quan trọng như số lượng người, sự kiện, theo dõi người lẫn trong đám đông, ...

Những khu vực đông người như ga tàu điện ngầm, trung tâm thương mại, quảng trường, sân vận động, ... thường cần 1 hệ thống kiểm soát đám đông. Khi số lượng người tăng đến 1 mức nào đó, dựa trên thông tin số lượng trích xuất từ video của camera, hệ thống sẽ tự động báo với người kiểm soát an ninh nhằm đưa ra biện pháp tức thời để tránh những rủi ro đáng tiếc xảy ra. Đó là 1 ví dụ cho thấy ý nghĩa to lớn về thông tin số lượng người. Ngoài vấn đề an ninh, việc nắm bắt số lượng người phân bố ở các quang cảnh khác nhau giúp doanh nghiệp có thể tối ưu cách bố trí không gian nếu sử dụng tốt thông tin phân tích được (1).

Việc áp dụng các hệ thống phân tích đám đông tự động dường như chưa phổ biến, hay còn mới mẻ ở các nước đang phát triển. Thay vào đó việc giám sát đám đông được thực hiện 1 cách thủ công, tốn nhiều công sức và có thể nói là không hiệu quả nếu cần giám sát trên 1 diện rộng với nhiều camera quay hình cùng 1 lúc.

Đề tài này nhắm đến thông tin về số lượng đám đông vì nó có thể áp dụng vào bài toán giao thông. Giao thông ở Việt Nam vào những giờ cao điểm thường rất đông đúc, trạng thái kẹt xe gây nhiều khó khăn đến các phương tiện tham gia giao thông, ảnh hưởng không nhỏ đến xã hội nước ta hiện nay. Nếu hệ thống phân tích tự động số lượng người tham gia giao thông trên các tuyến đường chính được áp dụng, việc nắm bắt tình hình sẽ là dễ dàng và không khó để có giải pháp hiệu quả cho vấn đề đó.

1.2 Ứng dụng của đề tài

Hệ thống ước lượng số lượng người trong đám đông có thể ứng dụng thực tế:

- Ước lượng mật độ người tham gia giao thông tại các nút ngã ba, ngã tư.
- Kiểm soát lượng người ở siêu thị lớn, bến xe, sân bay để thiết kế cách bố trí sắp xếp không gian hợp lý, tránh những điểm quá đông người và những điểm khác thì quá thưa thớt.
- Kiểm soát an ninh những điểm thường tụ tập đông người gây mất trật tự, khi lượng người tăng đến số lượng nhất định sẽ tự động thông báo đến người có trách nhiệm.

1.3 Hướng tiếp cận

Có 2 phương pháp để ước lượng số lượng người trong đám đông. Đó là phương pháp ước lượng trực tiếp và phương pháp ước lượng gián tiếp.

Ước lượng trực tiếp được thực hiện bằng cách nhận diện tất cả các đối tượng người trong ảnh. Số lượng các đối tượng nhận diện được chính là số lượng người trong ảnh. Phương pháp này hoạt động hiệu quả khi số lượng người trong ảnh không quá lớn, các đặc trưng người là rõ ràng đủ để hệ thống nhận diện chính xác. Với cách tiếp cận này cho ảnh thưa người, còn có thêm thông tin hữu ích khác là sự phân bố người trên vùng camera quan sát và hướng di chuyển của đối tượng trong video. Việc áp dụng phương pháp trực tiếp cho ảnh đặc biệt đông người, mà ở đó có sự che phủ lẫn nhau nhiều, các đặc trưng nhận diện đối tượng người bị che lấp. Kết quả dẫn đến là sự sai lệch không nhận diện được đối tượng người. Vì vậy, phương pháp ước lượng trực tiếp không đáng tin cậy khi áp dụng cho ảnh đông người.

Ước lượng gián tiếp được thực hiện bằng cách trích xuất đặc trưng từ ảnh. Đặc trưng tương ứng với số lượng người. Số lượng người khác nhau thì đặc trưng khác nhau. Ước lượng số lượng người theo hướng gián tiếp tin cậy hơn hướng tiếp cận trực tiếp, độ chính xác cao và giải quyết được phần nào vấn đề do sự che phủ người trong đám đông gây ra. Đề tài này được nghiên cứu dựa trên các phương pháp theo hướng tiếp cận gián tiếp nhằm đạt được độ chính xác cao khi ước lượng.

2. KIẾN THỨC NỀN TẢNG

2.1 Cân bằng sáng Histogram

Histogram (1) của 1 ảnh với L mức sáng từ $[0, L-1]$ là 1 hàm rời rạc ánh xạ số lần xuất hiện của mức sáng. Biểu diễn dưới dạng công thức:

$$h(r_k) = n_k$$

r_k – mức sáng thứ k , $k = 0, 1, 2, \dots, L-1$

n_k – số pixel trong ảnh có mức sáng r_k

Histogram được chuẩn hóa bằng cách chia cho tổng số pixel của ảnh. Biểu diễn dưới dạng công thức:

$$p(r_k) = n_k/MN$$

$p(r_k)$ – xác suất xuất hiện của mức sáng thứ k trong ảnh

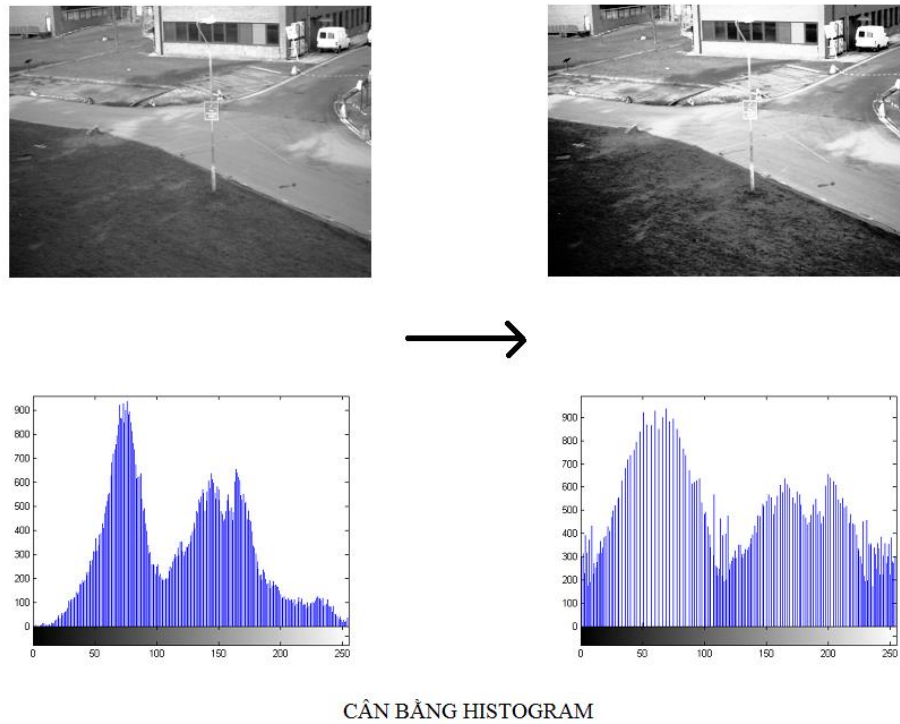
MN – tổng số pixel trong ảnh kích thước M hàng, N cột

Cân bằng sáng Histogram dùng để cân bằng các xác suất xuất hiện của mức sáng. Công thức biến đổi cân bằng sáng:

$$s_k = T(r_k) = (L-1) \sum_{j=0}^k p(r_j)$$

s_k – ánh xạ của r_k qua phép biến đổi

Giá trị s_k được làm tròn về số nguyên. Nếu có nhiều hơn 1 mức sáng trong ảnh đầu vào chiếu sang ảnh đầu ra vào cùng 1 mức sáng s_t thì histogram của ảnh đầu ra tại s_t sẽ là cộng dồn các xác suất xuất hiện tại các mức sáng ảnh input.

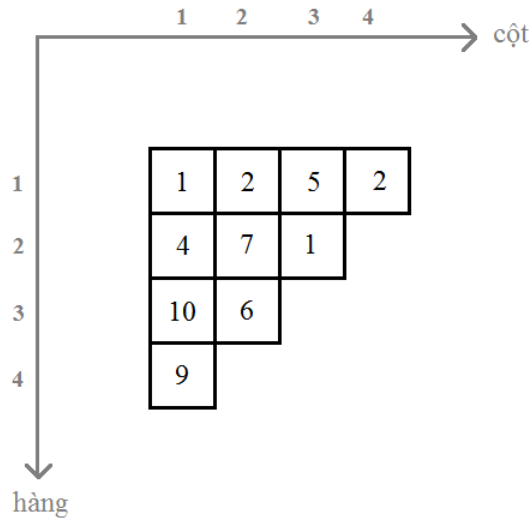


Hình 1. Ảnh, histogram trước và sau khi cân bằng sáng

2.2 Convolution

Tương quan (correlation) giữa mặt nạ và ảnh cho ra ảnh kết quả giữ lại các tính chất mong muốn khi thiết kế mặt nạ. Tương quan mặt nạ và ảnh có thể dùng để lọc các cạnh (mặt nạ Sobel) của ảnh, làm mờ (mặt nạ Gauss) ảnh, ...

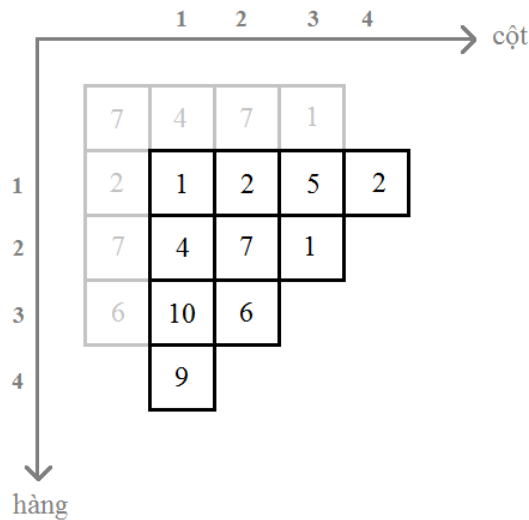
Cách tính toán tương quan giữa mặt nạ và ảnh:



Hình 2. Các pixel góc phía trên bên trái của ảnh

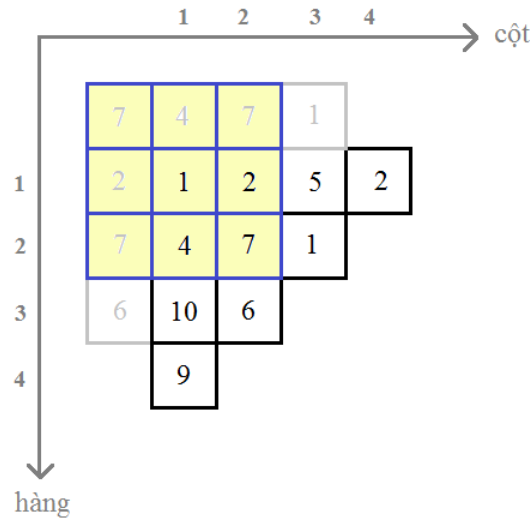
Giả sử ta có ảnh như hình 2 và mặt nạ kích thước 3x3 như sau:

$$S = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$



Hình 3. Xây dựng pixel nằm ngoài ảnh

Để có tính toán ảnh kết quả ta đặt vị trí mặt nạ lần lượt với các pixel của ảnh sao cho điểm giữa của mặt nạ ứng với pixel cần tính của ảnh. Các phần tử trong mặt nạ nhân tương ứng vị trí với giá trị của những pixel trong ảnh sau đó cộng chúng lại.



Hình 4. Vùng mặt nạ nhân tương ứng

Hình 4 minh họa cách tính cho pixel tọa độ (1,1). Với mặt nạ S như đã nêu trên thì giá trị tại điểm (1,1) của ảnh sau khi tương quan sẽ bằng $1*7 + 2*4 + 3*7 + 4*2 + 5*1 + 6*2 + 7*7 + 8*4 + 9*7 = 205$.

Ảnh đầu vào kích thước $M \times N$ thì ảnh kết quả sau khi tương quan cũng có kích thước $M \times N$. Mặt nạ chạy $M \times N$ lần tính toán như nêu trên.

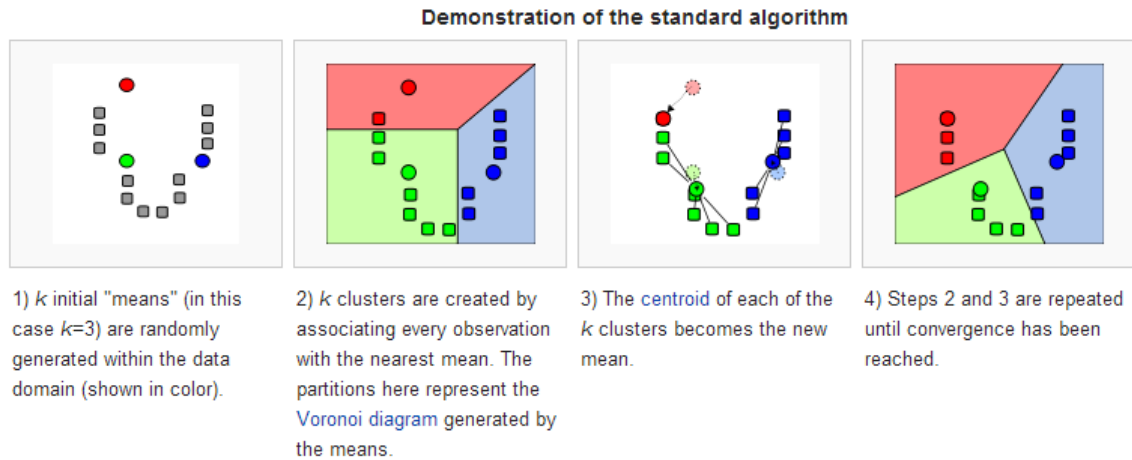
Cách tính toán convolution là xoay mặt nạ 180° sau đó thực hiện hiện tính toán tương quan.

2.3 Phân cụm K-means

K-means chia vùng n quan sát thành k cụm mà mỗi quan sát thuộc về cụm có giá trị mean gần nhất (2). Độ phức tạp tính toán K-means là NP.

K hạt ban đầu được gieo là K điểm ngẫu nhiên trong n quan sát cần phân cụm. Sau đó n quan sát được tính toán phân thành K cụm, tiêu chí xét quan sát thuộc cụm nào dựa trên khoảng cách từ nó đến mean của cụm, khoảng cách nhỏ nhất được chọn. Điểm mean của các cụm được tính toán lại. Quá trình như thế được lặp đi lặp lại cho đến khi

hội tụ (các điểm mean của các cụm không thay đổi) hoặc đạt đến số lần lặp nhất định.



Hình 5. Minh họa giải thuật K-means

2.4 Gradient

Đạo hàm ảnh theo trục x tương ứng với phép convolution mặt nạ Sobel với ảnh. Đạo hàm ảnh theo trục y tương ứng với phép convolution chuyển vị của mặt nạ Sobel với ảnh.

Mặt nạ Sobel:

$$S_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Mặt nạ Sobel chuyển vị:

$$S_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Ảnh gradient là sự thay đổi có hướng về mức sáng / màu sắc trong ảnh. Ảnh gradient được tạo ra bằng cách convolution mặt nạ Sobel theo trục x / trục y với ảnh. Để ảnh gradient thể hiện sự biến đổi mức sáng theo cả 2 chiều x và y:

$$G(x, y) = \sqrt{S_x^2(x, y) + S_y^2(x, y)}$$

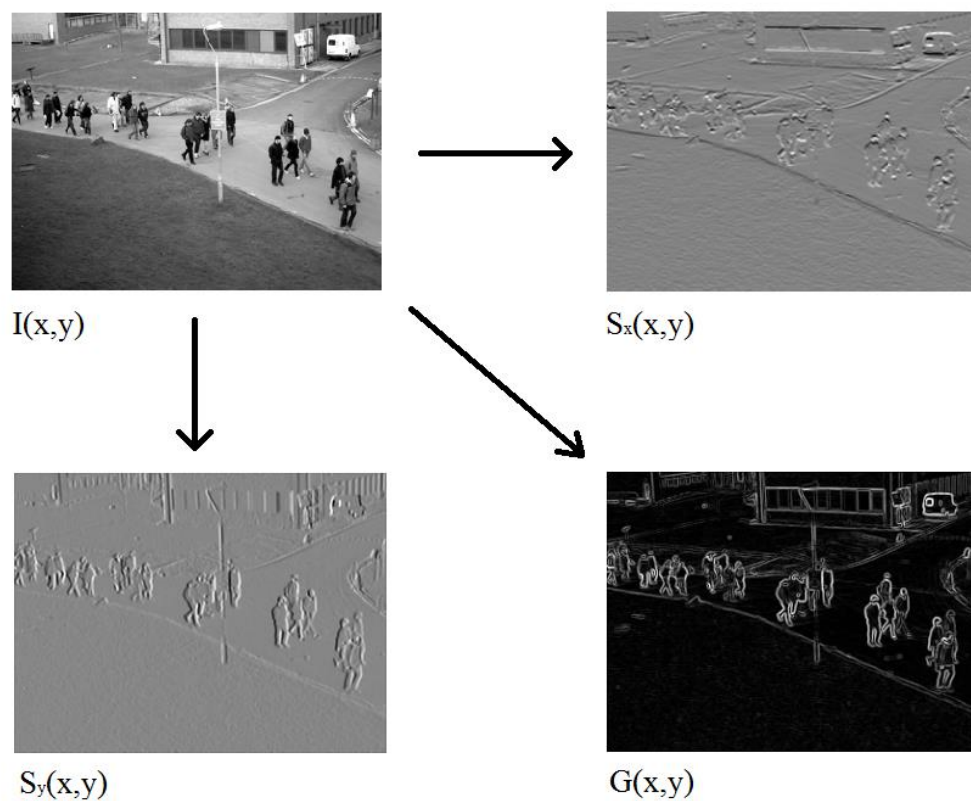
$I(x, y)$ - ảnh gốc

• - phép convolution

$$S_x(x,y) = I(x,y) \cdot S_x$$

$$S_y(x,y) = I(x,y) \cdot S_y$$

$G(x,y)$ - ảnh gradient theo cả 2 chiều x và y



Hình 6. Ảnh gradient

3. Phương pháp ước lượng

3.1 Tổng quan các phương pháp khác

3.1.1 Phương pháp sử dụng hình dạng của nhóm người

Phương pháp này xây dựng trước 1 tập các hình dạng bên ngoài khi 1 nhóm số lượng nhất định người đi có thể tạo ra, ước lượng dựa trên trích xuất hình dạng người/nhóm người trong ảnh để tra cứu tương ứng trong dữ liệu đã xây dựng (1). Ảnh so sánh với background (không có người) để lọc ra được foreground. Ảnh foreground tiếp theo sẽ được biến đổi Fourier sang miền tần số, lọc giữ lại những tần số thấp và tái xây dựng lại.

Số lượng người tương ứng với hình dạng đã được xây dựng trước được index vào Look-up Table để tăng tốc độ tìm kiếm hình dạng gần giống khi ước lượng. Phương pháp này không nêu rõ số lượng tập dữ liệu xây dựng sẵn bao nhiêu là đủ. Độ tin cậy của phương pháp này cũng chưa chắc chắn.

3.1.2 Phương pháp thống kê đặc trưng góc chuyển động

Phương pháp này ước lượng số lượng người dựa trên các đặc trưng về góc chuyển động có trong video được nghiên cứu trong (2). Với mỗi frame ảnh thực hiện tính ảnh gradient, dựa trên ảnh gradient tính được ma trận covariance. Bộ giá trị eigenvalue của ma trận covariance phản ánh đặc điểm của pixel đó. Nếu 1 eigenvalue lớn và eigenvalue còn lại nhỏ thì đó là pixel cạnh. Nếu cả 2 giá trị eigenvalue đều nhỏ thì pixel đó thuộc vùng trơn (không có góc, cạnh). Các điều kiện được xây dựng để xem xét cặp eigenvalue đó có được xem là góc hay không được nhóm tác giả đưa ra.

Những pixel là góc được xem là “có chuyển động” nếu so với frame trước vị trí của nó có thay đổi. Với vị trí tĩnh camera, cảnh vật tĩnh thì giữa 2 khung hình các pixel đặc trưng cho góc không đổi. Vì vậy sẽ loại được những đặc trưng đứng yên, giữ được các đặc trưng góc có chuyển động. Số lượng người được ước lượng bằng cách lấy số lượng các đặc trưng góc chuyển động trong khung hình chia cho số đặc trưng góc trung bình của 1 người

$$N_{\text{people}} = N_c / N_1$$

N_{people} – số lượng người có trong khung hình

N_c – số lượng đặc trưng góc chuyển động có trong khung hình

N_1 – số lượng đặc trưng góc trung bình của 1 người

3.1.3 Phương pháp sử dụng đặc trưng SURF

Phương pháp này sử dụng đặc trưng SURF trích xuất từ ảnh và mô hình học máy ϵ -SVR (3). Các đặc trưng SURF được tính toán trong các khung hình. Tương tự phương pháp trên, chỉ những đặc trưng SURF chuyển động được giữ lại. Đặc trưng SURF còn lại được tiến hành phân cụm theo giải thuật Graph-Based Clustering, ý tưởng cơ bản của cách phân cụm này là xây dựng cây phủ tối thiểu của các điểm đặc

trung SURF. Mỗi nút (node) trên cây là pixel đặc trưng, cạnh (edge) của cây nối 2 nút có trọng số là khoảng cách giữa 2 pixel được nối. Các điều kiện được xây dựng để xóa các cạnh có trọng số lớn (khoảng cách giữa 2 pixel xa). Khi đó các pixel được nối liền với nhau bởi các cạnh (liên thông) được phân thành 1 nhóm. Những nhóm có ít pixel đặc trưng bị loại bỏ.

Biên của cụm các pixel được xây dựng tạo thành diện tích hình chữ nhật. Khoảng cách từ cụm trong ảnh đến camera được ước lượng theo Inverse Perspective Mapping (IPS). Mô hình học máy hồi quy ϵ - SVR được chọn để huấn luyện tập dữ liệu đã được gán nhãn tương ứng với số người.

$$n_{\text{people}} = f(n_{\text{points}}, p, d)$$

n_{people} – số lượng người

f – hàm máy học

n_{points} – số lượng điểm đặc trưng SURF

p – mật độ điểm đặc trưng SURF trên diện tích cụm

d – khoảng cách IPS

3.1.4 Nhận xét

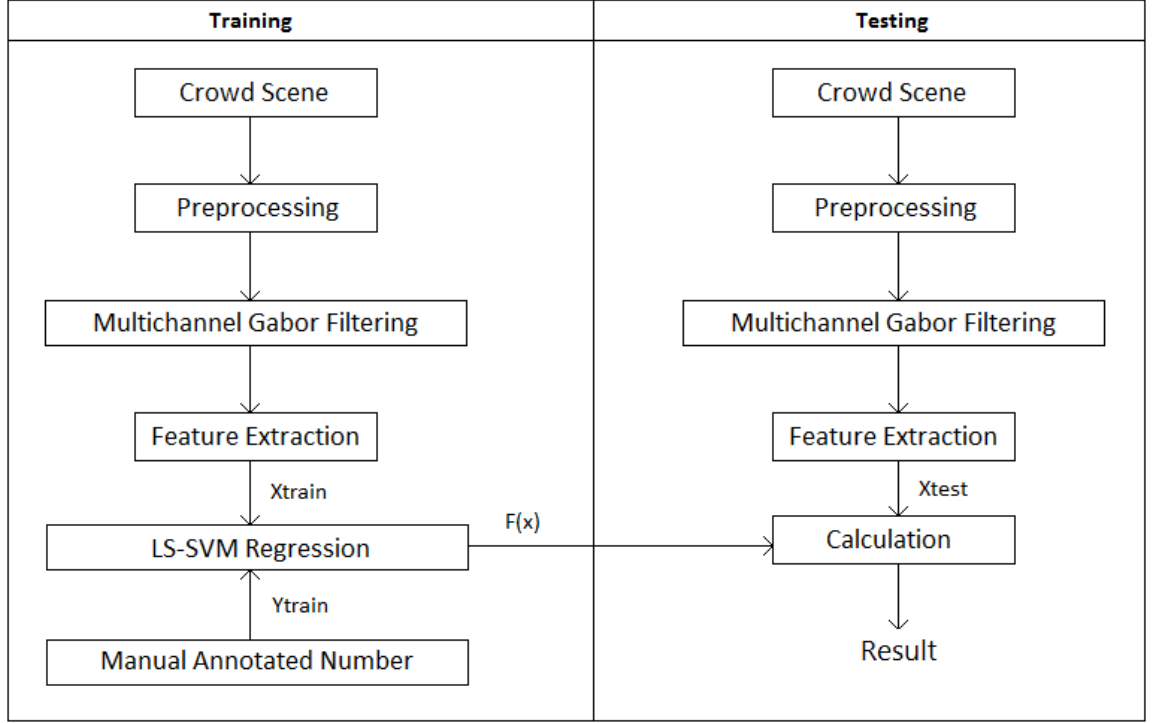
Nhìn chung các phương pháp trên độ chính xác còn chưa cao. Một số bước trong các phương pháp còn chưa đáng tin cậy như:

- Số lượng tập dữ liệu hình dạng người/nhóm (phương pháp 3.1.1)
- Hình dạng nhóm người giống nhau nhưng số lượng người khác nhau do che phủ (phương pháp 3.1.1)
- Số đặc trưng góc của 1 người không phải lúc nào cũng giống nhau ở mọi góc nhìn, khoảng cách (phương pháp 3.1.2)
- Chia cụm bằng cách lược bỏ cạnh lớn trên cây phủ tối thiểu không phải lúc nào cũng đảm bảo chính xác (phương pháp 3.1.3)

3.2 Phương pháp Gabor filter

3.2.1 Tổng quan phương pháp

Phương pháp này được nhóm tác giả đi trước nghiên cứu (6). Hệ thống ước lượng gồm 2 phần: huấn luyện (train) và kiểm tra (test). Những ảnh được huấn luyện được trích xuất đặc trưng về kết cấu ảnh bằng bộ lọc Gabor. Việc đưa các bộ đặc trưng và số lượng người tương ứng vào mô hình học máy sẽ cho ra kết quả là 1 hàm toán học thể hiện phép chiếu từ không gian đặc trưng sang không gian số lượng người. Giai đoạn kiểm tra sử dụng phép chiếu có được từ giai đoạn huấn luyện để thực hiện đánh giá kết quả độ chính xác.



Hình 7. Kiến trúc hệ thống phương pháp Gabor filter

Ảnh được xử lý thu giảm về kích thước 320x240 pixel và cân bằng histogram trước khi đi qua bộ lọc Gabor. Cân bằng histogram nhằm mục đích giảm bớt ảnh hưởng của chiều sáng.

Ảnh sau khi lọc với Gabor có kích thước bằng ảnh trước khi lọc. Đặc trưng của ảnh là mean và variance trên toàn bộ ảnh sau khi lọc. Mô hình học máy trong phần huấn luyện được chọn sử dụng là Least Squares Support Vector Machine (LS-SVM).

3.2.2 Thiết kế bộ lọc Gabor

Công thức thiết lập bộ lọc Gabor chẵn và lẻ:

$$h_e(x, y) = g(x, y) \cos(2\pi\omega(x\cos\theta + y\sin\theta))$$

$$h_o(x, y) = g(x, y) \sin(2\pi\omega(x\cos\theta + y\sin\theta))$$

$$g(x, y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

$h_e(x, y)$ – bộ lọc Gabor chẵn

$h_o(x, y)$ – bộ lọc Gabor lẻ

$g(x, y)$ – hàm Gauss 2-D

Bộ lọc Gabor ở dạng phức $h(x, y) = h_e(x, y) + jh_o(x, y)$ (7)

Do bộ lọc Gabor ta chọn ở miền quang phổ nên chỉ sử dụng phần thực (Gabor chẵn $h_e(x,y)$ hoặc phần ảo (Gabor lẻ $h_o(x,y)$) để xây dựng bộ lọc.

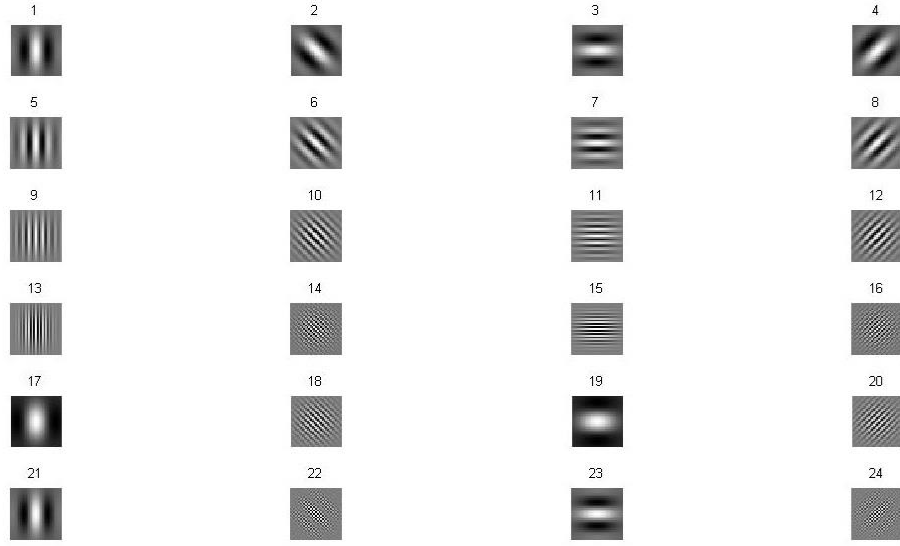
Các tham số ω , σ , θ được thiết lập như sau:

$$\omega = 2, 4, 8, 16, 32, 64$$

$$\sigma = 1/\omega$$

$$\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$$

Tần số trung tâm ω , góc (hướng) θ và hằng số quang phổ σ là 3 tham số quan trọng thiết lập kênh Gabor. Với 6 tần số trung tâm ω và 4 góc θ ta xây dựng được 24 kênh Gabor. Kích thước bộ lọc Gabor trên mỗi kênh tương ứng là 31x31 pixel.



Hình 8. Hình ảnh 24 bộ lọc Gabor (chẵn)

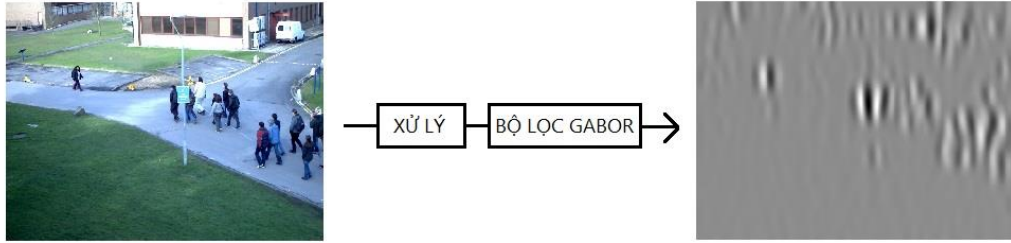
Ảnh đầu ra của bộ lọc là convolution giữa ảnh đầu vào và bộ lọc Gabor

$$O_e(x, y) = I(x, y) \bullet h_e(x, y)$$

$$O_o(x, y) = I(x, y) \bullet h_o(x, y)$$

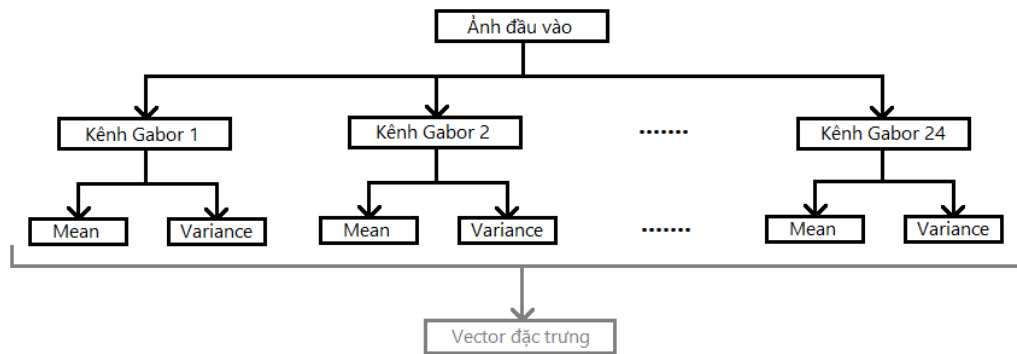
$$O(x, y) = \sqrt{O_e^2(x, y) + O_o^2(x, y)}$$

Nếu sử dụng bộ lọc Gabor chẵn $h_e(x,y)$ sẽ thu được ảnh đầu ra tương ứng. Tương tự cho Gabor lẻ $h_o(x,y)$. Ảnh đầu ra $O(x,y)$ được tính toán dựa trên ảnh đầu ra sau khi bộ lọc chẵn và lẻ như công thức nêu trên. Việc sử dụng ảnh đầu ra nào để trích đặc trưng có ảnh hưởng đến kết quả độ chính xác. Theo các kết quả thí nghiệm thì bộ lọc Gabor lẻ $h_o(x,y)$ là hiệu quả nhất.



Hình 9. Ảnh sau khi đi qua kênh Gabor thứ nhất (lẽ)

Ảnh sau khi xử lý thu giảm kích thước và cân bằng histogram, sẽ đi qua bộ lọc 24 bộ lọc Gabor thu được 24 ảnh đầu ra. Mỗi ảnh đầu ra tính mean và variance, với 24 ảnh đầu ra ta có 24 bộ giá trị mean và variance. Nối các bộ giá trị này với nhau ta được vector đặc trưng có chiều dài 48.



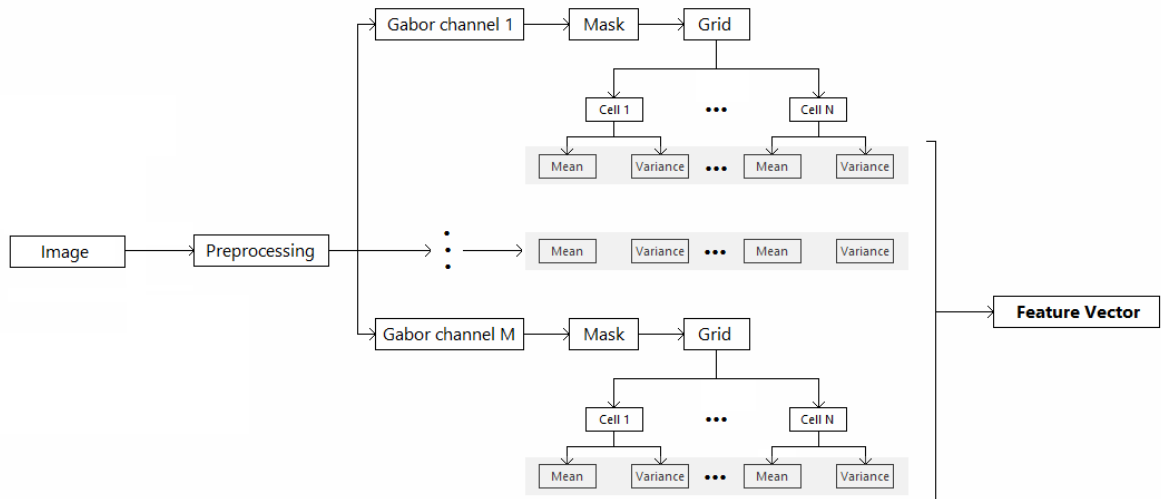
Hình 10. Trích đặc trưng theo mean và variance của các ảnh đầu ra

Phần cuối của giai đoạn huấn luyện là đưa cặp giá trị (x_{train} , y_{train}) vào huấn luyện cho mô hình học máy LS-SVM. Vector x_{train} có 48 phần tử tương ứng với ảnh có nhãn y_{train} người.

Giai đoạn kiểm tra ta tính toán đặc trưng của ảnh tương tự như giai đoạn train. Giá trị vector x_{test} (48 phần tử) sẽ được mô hình học máy tính toán và trả về kết quả $result$. Độ chính xác của $result$ phụ thuộc nhiều vào cách trích đặc trưng.

3.3 Phương pháp 1 - Chia lưới

Phương pháp chia lưới vẫn sử dụng các bộ lọc Gabor, thay đổi cách trích xuất đặc trưng từ chỉ đơn thuần lấy mean và variance sau khi lọc Gabor sang áp dụng mặt nạ (mask) và lưới rồi sau đó lấy mean và variance trên mỗi ô lưới (cell).



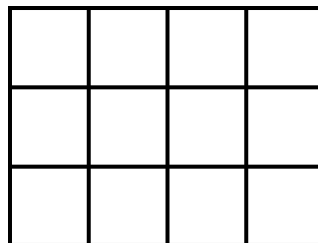
Hình 11. Phương pháp chia lưới

Ta phân tách những vật thể như nhà, cây cối, hoặc địa hình không thể đi được với chỗ có thể di chuyển được như trên đường, bãi cỏ, ... được vùng quan sát. Những pixel nằm ngoài vùng quan sát sẽ bị loại bỏ.



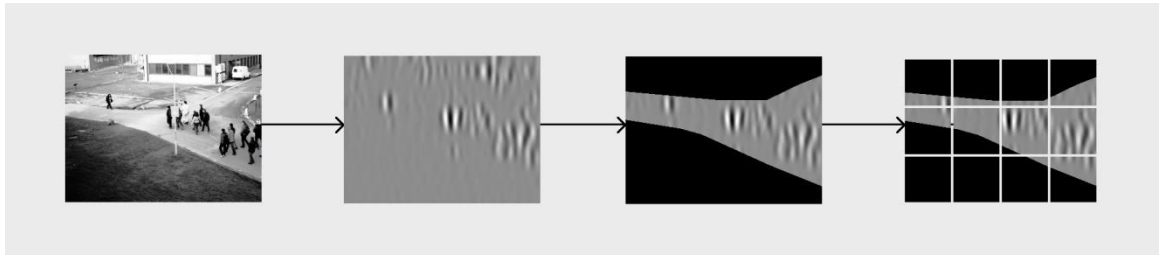
Hình 12. Thiết lập mặt nạ

Lưới được chọn với số lượng ô không quá lớn. Mỗi ô của lưới chứa các thành phần của ảnh, khi đó ô sẽ mang tính chất cục bộ.



Hình 13. Lưới 3x4

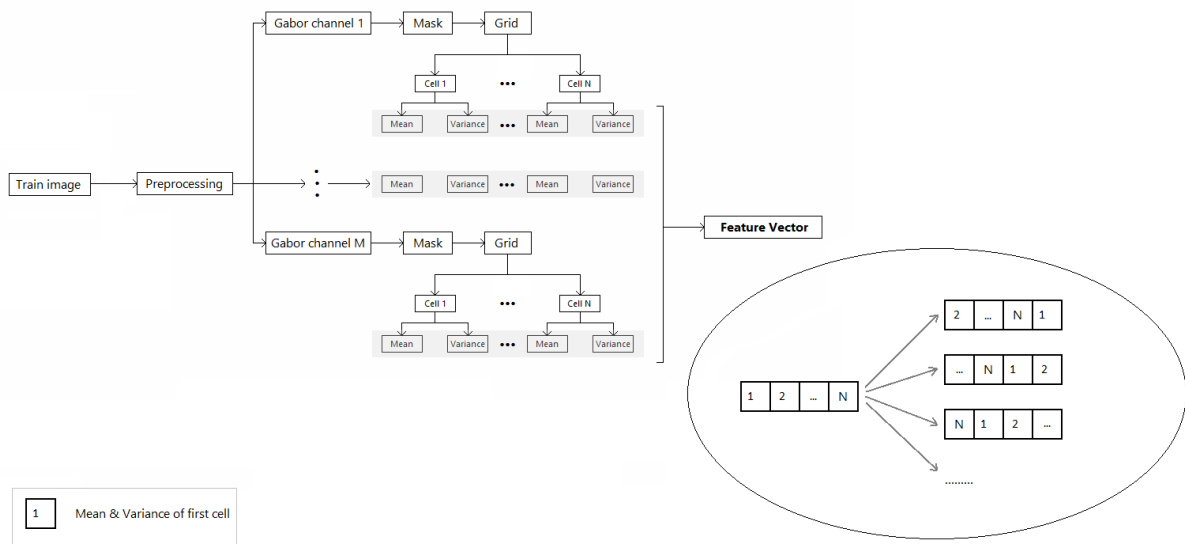
Vector đặc trưng được xây dựng bằng cách ghép mean và variance của tất cả các ô trên tất cả các kênh Gabor. Nếu số kênh Gabor là M , lưới có N ô thì chiều dài của vector đặc trưng của 1 ảnh bằng $2 \times N \times M$.



Hình 14. Áp dụng chia lưới

3.4 Phương pháp 2 - Hoán vị

Phương pháp này mở rộng hơn so với phương pháp chia lưới bằng cách huấn luyện các hoán vị vị trí của ô trên lưới. Khi đó số lượng tập huấn luyện tăng lên gấp nhiều lần.

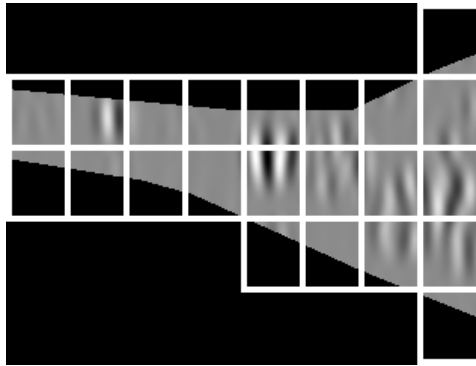


Hình 15. Phương pháp hoán vị

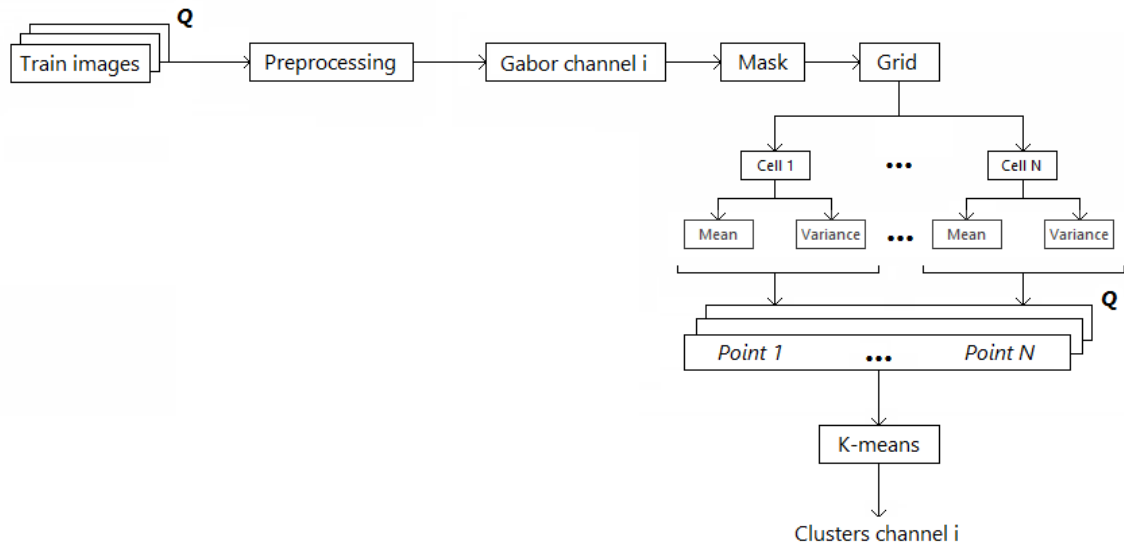
Hoán vị các cặp mean và variance ở kênh thứ nhất. Vị trí tương đối giữa các cặp mean và variance ở tất cả các kênh đều là giống nhau. Phương pháp hoán vị này nhằm tăng tính tổng quát của phương pháp chia lưới. Hoán vị ở tất cả các vị trí để bao hàm tất cả vị trí người phân bố trên các ô lưới với cùng 1 lượng người nhất định.

3.5 Phương pháp 3 - Bag of words

Phương pháp ứng dụng Bag of words tương tự với phương pháp chia lưới đã trình bày trước đó nhưng khác về mặt trích xuất đặc trưng. Phương pháp chia lưới sử dụng mean và variance của ô lưới trên 24 kênh ghép lại để tạo vector đặc trưng chiều dài $24 \times 2 \times N$ (với N là số ô lưới). Do ta có sử dụng mặt nạ để loại bỏ vùng nằm ngoài quan sát, vì vậy những ô lưới trên vùng đó là vô nghĩa. Vì vậy, ta chỉ lấy những ô nằm trong vùng quan sát để tính mean và variance. Những ô vô nghĩa sẽ bị lược bỏ không sử dụng tới.

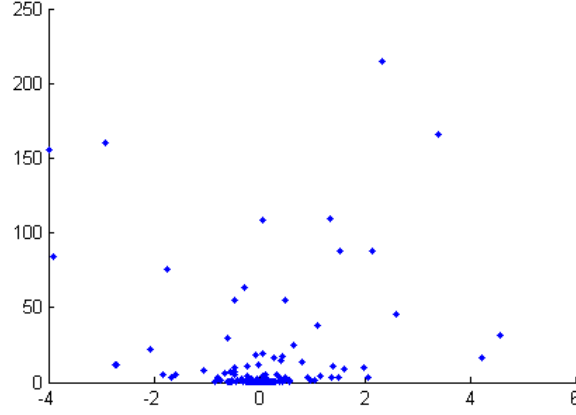


Hình 16. Lưới trên vùng quan sát - phương pháp Bag of words

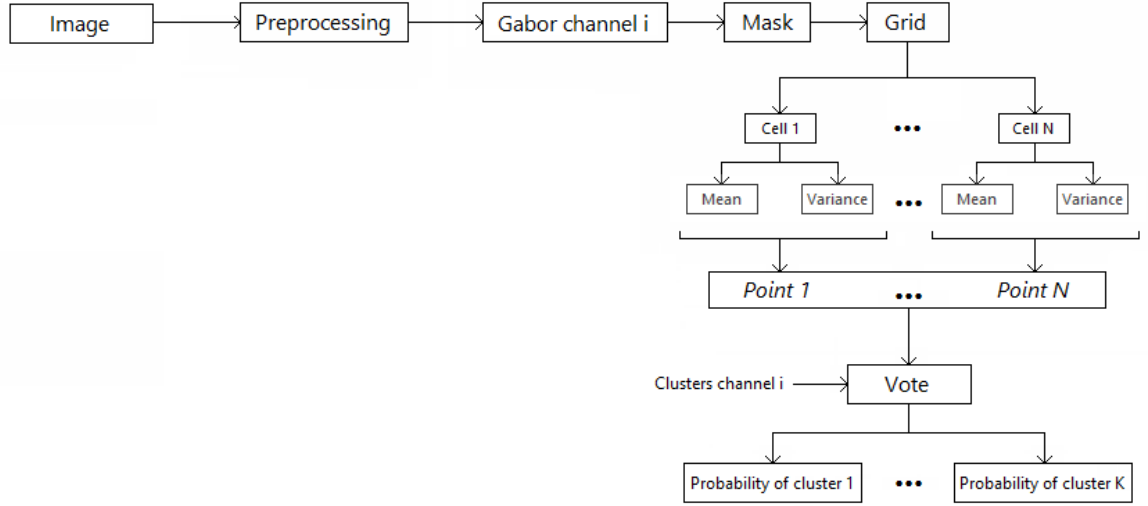


Hình 17. Phân cụm - phương pháp Bag of words

Sau khi tính được các cặp giá trị mean và variance trên mỗi ô của lưới (phương pháp chia lưới), xem mỗi cặp mean và variance của ô là 1 điểm trên không gian 2 chiều. Với Q ảnh train, lưới N ô thì trên mỗi kênh Gabor ta sẽ có $Q \times N$ điểm. Phân cụm $Q \times N$ điểm này thành K cụm.



Hình 18. Các điểm trong hệ trục mean, variance



Hình 19. Trích xuất đặc trưng trên mỗi kênh - phương pháp Bag of words

Sau khi phân cụm trên mỗi kênh thì ta có được điểm trung tâm của mỗi cụm. Sự phân thành K cụm sẽ cho ra K điểm trung tâm. Cùng cách làm như trên với ảnh đầu vào trên mỗi kênh, lưới kích thước N, sẽ có N điểm trên hệ trục 2 chiều mean và variance. N điểm này được thống kê vào K cụm đã chia ở bước trước đó bằng cách so sánh khoảng cách từ điểm cần xét đến tất cả các điểm trung tâm của cụm, khoảng cách nhỏ nhất sẽ được chọn. Khi 1 điểm được xác định thuộc 1 cụm nào đó (gần điểm trung tâm cụm đó nhất), giá trị thống kê của cụm đó tăng lên 1. Sau khi thống kê cả N điểm vào K cụm, bước tiếp theo ta chuẩn hóa bằng cách chia giá trị thống kê của mỗi cụm cho tổng số điểm.

$$Cluster\ i_n = \frac{Cluster\ i}{N}$$

Cluster i_n – giá trị thống kê của cụm thứ i đã được chuẩn hóa

Cluster i – giá trị thống kê của cụm thứ i, số điểm thuộc cụm thứ i

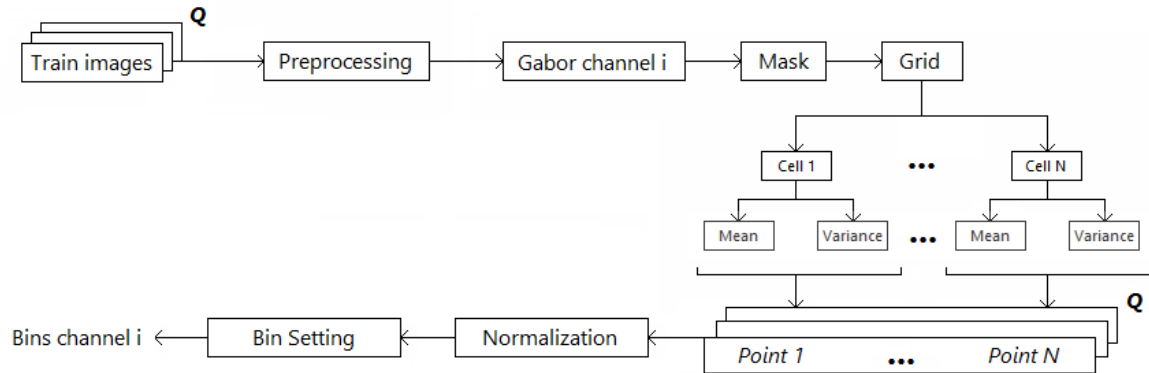
N – tổng số điểm được thống kê

Trên mỗi kênh Gabor ta thu được K giá trị thống kê, với 24 kênh Gabor thì chiều dài vector đặc trưng là $24 \times K$. Phần còn lại của phương pháp ta chỉ việc đưa vector đặc trưng $24 \times K$ vào mô hình học máy LS-SVM. Giai đoạn kiểm tra tương tự, trích xuất đặc trưng thành vector $24 \times K$ đưa vào mô hình học máy để nhận kết quả ước lượng.

Một hạn chế của phương pháp Bag of words là giải thuật phân cụm K-means khá phụ thuộc vào việc gieo hạt giống (seed) ban đầu. Gieo hạt giống là ngẫu nhiên, nên kết quả các điểm thuộc cùng 1 cụm chỉ đảm bảo tương đối là chúng có tính chất gần giống nhau (mean và variance).

3.6 Phương pháp 4 - Histogram

Phương pháp Histogram trích xuất đặc trưng bằng cách thống kê các điểm theo các bin. Phương pháp vẫn sử dụng 24 kênh Gabor, mặt nạ, lưới (chia lưới theo phương pháp Bag of words) và xem các điểm trên mỗi ô lưới là điểm trên không gian 2 chiều mean và variance như ở phương pháp Bag of words. Điểm thay đổi là thay vì phân cụm K-means, phương pháp này sẽ phân cụm theo bin.



Hình 20. Phân chia bin - phương pháp Histogram

Như sơ đồ minh họa ở trên, với Q ảnh train trên mỗi kênh Gabor, lưới N ô thì có $Q \times N$ điểm trên hệ trục 2 chiều. Mỗi điểm được chuẩn hóa bằng cách chia cho độ dài vector của nó.

$$(m_N, v_N) = \left(\frac{m}{\sqrt{m^2 + v^2}}, \frac{v}{\sqrt{m^2 + v^2}} \right)$$

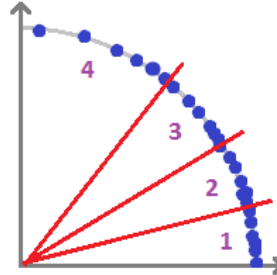
(m_N, v_N) – điểm sau khi chuẩn hóa

(m, v) – điểm trước khi chuẩn hóa

m – giá trị mean của ô

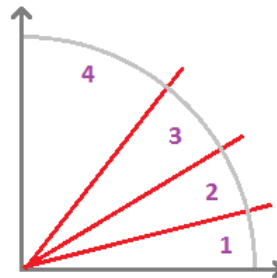
v – giá trị variance của ô

Điểm sau khi chuẩn hóa sẽ nằm trên cung đường tròn (độ dài vector từ gốc tọa độ đến điểm đều bằng 1). Ta thiết lập giá trị góc của mỗi bin sao cho mỗi bin đều chứa số lượng các điểm bằng nhau.



Hình 21. Xây dựng bin - phương pháp Histogram

Giả sử có 100 điểm trên cung tròn góc $\frac{1}{4}$ thứ nhất, số bin cần chia là 4 thì mỗi bin sẽ chứa 25 điểm. Góc của mỗi bin phụ thuộc vào vị trí các điểm phân bố trên cung tròn. Nếu ta chia bin theo các góc bằng nhau, các điểm phân bố không đồng đều vì thế có khả năng một số bin chứa rất nhiều điểm, một số bin còn lại chứa ít hoặc không chứa điểm nào. Do vậy, việc chia bin đồng đều về số lượng điểm là hợp lý.

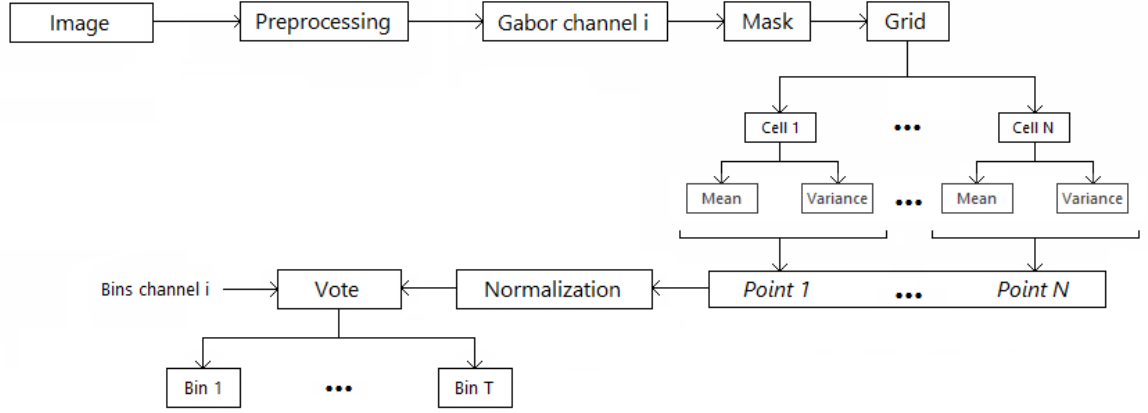


Hình 22. Bin tương ứng với góc - phương pháp Histogram

Điểm tọa độ $A(m_N, v_N)$, góc tạo bởi vector OA và trục hoành:

$$angle = \arctan\left(\frac{v_N}{m_N}\right)$$

Điểm được xem là thuộc bin nếu góc của điểm đó tạo ra so với trục hoành nằm trong góc thuộc bin đó. Sau khi các bin đã được thiết lập, ta tiến hành tính toán trích xuất vector đặc trưng.



Hình 23. Trích xuất đặc trưng trên mỗi kênh - phương pháp Histogram

Thực hiện thống kê các điểm sau khi được chuẩn hóa cho các bin. Giả sử sử lưới N ô, số bin là T . Một ảnh sẽ có N điểm trên hệ trục mean và variance, thống kê N điểm này xem có bao nhiêu điểm thuộc bin 1, bin 2, ... bin T . Với mỗi điểm khi được xác định thuộc bin j (với $1 \leq j \leq T$), có 2 cách để tăng giá trị thống kê cho bin j này. Cách thứ nhất tương tự phương pháp Bag of words, đó là tăng bin j thêm 1, sau khi thống kê cả N điểm, giá trị mỗi bin đem chia cho N ta nhận được xác suất xuất hiện của bin đó trong ảnh. Với cách thứ hai thì thay vì tăng bin j lên 1, ta tăng bin j thêm giá trị độ lớn của vector nối từ gốc tọa độ O đến điểm đó (chưa chuẩn hóa). Công thức bên dưới minh họa lần lượt 2 cách chuẩn hóa đã đề cập:

$$\text{Thống kê: Bin } j = \text{Bin } j + 1$$

$$\text{Chuẩn hóa: Bin } j_n = \frac{\text{Bin } j}{N}$$

$\text{Bin } j_n$ – giá trị thống kê của bin j đã chuẩn hóa

$\text{Bin } j$ – giá trị thống kê của bin j trước chuẩn hóa

N – tổng số điểm

$$\text{Thống kê: Bin } j = \text{Bin } j + \sqrt{m^2 + v^2}$$

$\text{Bin } j$ – giá trị thống kê của bin j

m – giá trị trên trục hoành (trục mean) của điểm tọa độ (m, v)

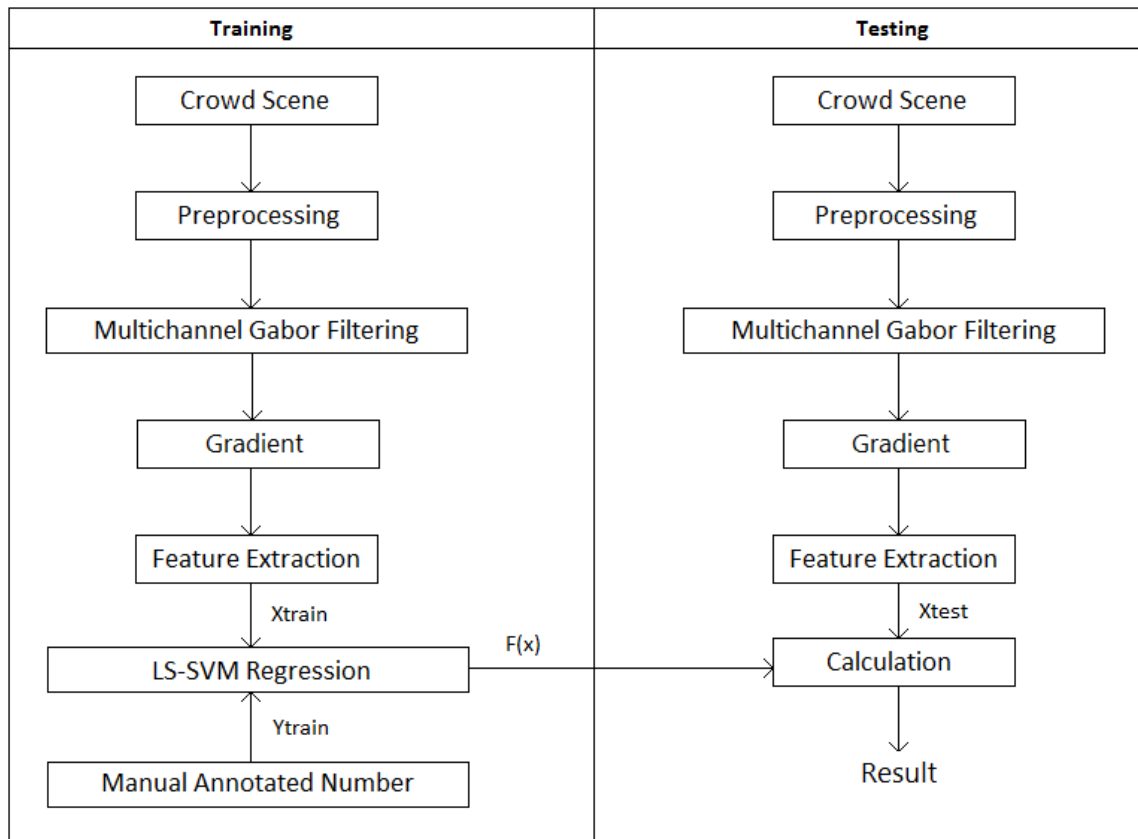
v – giá trị trên trục tung (trục variance) của điểm tọa độ (m, v)

Phương pháp sử dụng 24 kênh lọc Gabor, trên mỗi kênh Gabor thống kê được giá trị của T bin, do đó độ dài vector đặc trưng trích xuất được là $24 \times T$. Vector đặc trưng này được đưa vào mô hình học máy LS-SVM để huấn luyện, đối với ảnh huấn

luyện kèm nhãn số lượng người, hoặc để kiểm tra, đối với ảnh cần ước lượng số lượng người.

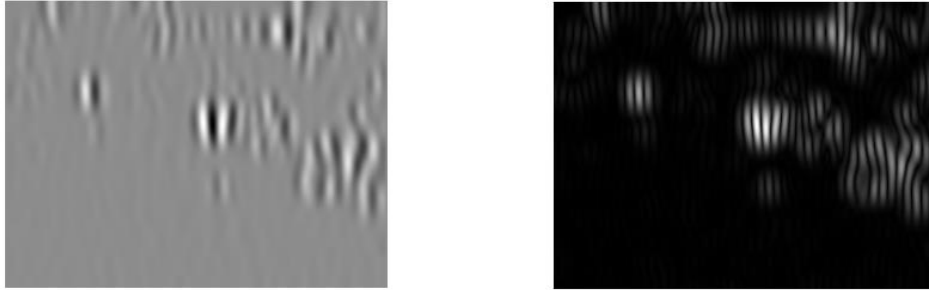
3.7 Phương pháp 5 - Gradient

Phương pháp Gradient sử dụng ảnh gradient của ảnh đầu ra của kênh Gabor. Tiếp theo, các bước trích đặc trưng tương tự các phương pháp đã trình bày nêu trên như gabor filter, chia lưới, bag of words, histogram.



Hình 24. Phương pháp Gradient

Ảnh sau khi lọc với Gabor sẽ lấy các đường nét theo hướng phụ thuộc vào góc θ . Ảnh gradient của nó sẽ làm nổi bật các đường nét này, cụ thể là ở những vùng ít biến đổi ảnh gradient của nó (theo cả 2 chiều x và y) sẽ có giá trị thấp gần với 0, những vùng có biến đổi mức sáng sẽ có giá trị lớn tương ứng mức độ biến đổi.



Hình 25. Ảnh lọc Gabor (trái) và ảnh gradient tương ứng (phải)

4. KẾT QUẢ THÍ NGHIỆM

4.1 Dataset

Dữ liệu hình ảnh lấy từ tập PETS 2009 (8). Bộ dữ liệu bao gồm S0, S1, S2, S3, mỗi tập dữ liệu được dùng với mục đích khác nhau. Tập S0 chứa các hình nền của cả 8 góc quay camera. Tập S1 phục vụ cho các thí nghiệm về đếm số lượng người và ước lượng mật độ. Tập S2 phục vụ cho các thí nghiệm về theo dõi người. Tập S3 phục vụ cho các thí nghiệm về phân tích luồng di chuyển và phát hiện sự kiện của đám đông.

Cấu trúc tập dữ liệu S1 gồm L1, L2 và L3 trên 4 góc quay camera View 1, View 2, View 3, View 4. Dữ liệu L1 là cảnh đám đông đi với mật độ trung bình, L2 là cảnh đám đông đi với mật độ dày đặc, L3 là cảnh đám chạy (mật độ trung bình) và ảnh hưởng của ánh sáng như chói sáng, bóng râm. Trong mỗi tập L có 2 thời gian quay khác nhau, mỗi thời gian quay đều chứa đầy đủ video 4 góc quay camera. Độ phân giải hình ảnh video mỗi camera là 768x576, xấp xỉ 7 khung hình/s.

Các thí nghiệm tôi thực hiện lấy từ bộ dữ liệu trong tập S1 trên cả L1, L2 và L3 với 3 góc quay camera View 1, View 2, View 3. Tổng số là 2260 ảnh. Cụ thể như sau:

1. S1-L1-1-V1
2. S1-L1-1-V2
3. S1-L1-1-V3
4. S1-L1-2-V1
5. S1-L1-2-V2
6. S1-L1-2-V3
7. S1-L2-1-V1
8. S1-L2-1-V2
9. S1-L2-1-V3
10. S1-L3-1-V1
11. S1-L3-1-V2
12. S1-L3-1-V3



View 1



View 2



View 3

Hình 26. Hình ảnh trong tập dataset

Các tập thí nghiệm nêu trên được chọn để đảm bảo đám đông di chuyển trong vùng mặt nạ, tạo sự thống nhất và công bằng khi so sánh độ chính xác giữa các phương pháp.

4.2 Kết quả so sánh

Chương trình được hiện thực trên MATLAB R2012b (8.0.0.783). Máy tính thực thi chương trình có cấu hình Intel(R) Core(TM) i5-3337U 1.80GHz, RAM 4GB, chạy trên hệ điều hành Windows 8.1 Pro.

Mô hình học máy LS-SVM được tham khảo từ bộ thư viện (8).

Thiết lập lưới của phương pháp chia lưới và hoán vị kích thước khoảng 2x3, 3x4. Số lượng ô lưới trong 2 phương pháp này không được quá nhiều.

Thiết lập lưới của phương pháp Bag of words và Histogram kích thước khoảng 12x32, 16x32, 20x40 và số cụm/bin khoảng từ 6 đến 10 là thích hợp.

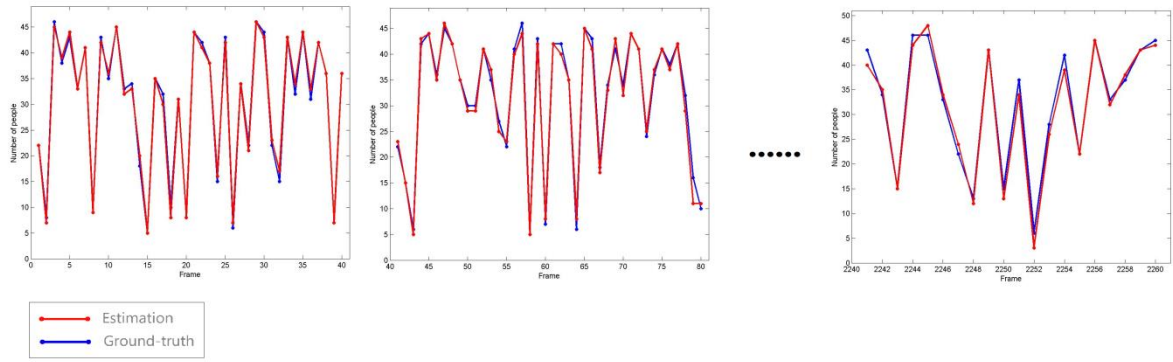
Tập kiểm tra	Số ảnh	Người			
		Tối thiểu	Trung bình	Tối đa	Variance
Data1 (gồm 12 video, 3 góc quay camera)	2260	1	30.74	47	136.05
Data2 (S1-L1-1-V1)	220	5	22.76	35	66.19
Data3 (S1-L3-1-V1)	90	7	24.47	40	122.9

Bảng 1. Thông tin dữ liệu thí nghiệm

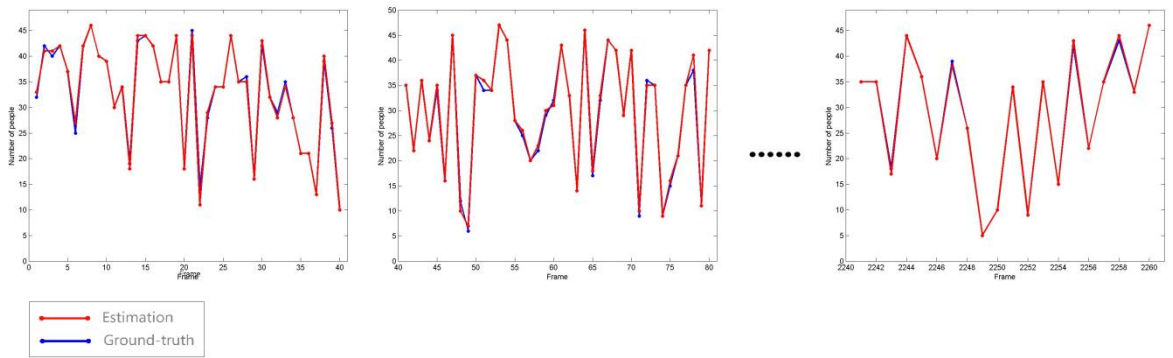
Kiểm tra trong các thí nghiệm là kiểm tra K-fold. Bộ lọc Gabor lẻ là hiệu quả nhất, nên ở tất cả các thí nghiệm đều sử dụng bộ lọc này. Số K để thực hiện kiểm tra K-fold được chọn bằng 5. Kiểm K-fold thì số lượng ảnh trong mỗi fold bằng nhau, vì vậy một vài ảnh cuối cùng sẽ không được đem vào tập kiểm tra (số ảnh này nhỏ hơn 5).

Tập kiểm tra	Độ chính xác của phương pháp (%)				
	Gabor filter	Chia lưới	Hoán vị	Bag of words	Histogram
Data1	95.82	97.60	93.41	93.30	92.80
Data2	96.37	97.46	95.43	96.34	96.05
Data3	93.89	95.95	92.62	95.47	94.18

Bảng 2. Độ chính xác của các phương pháp trên Data1, Data2, Data3



Hình 27. Biểu đồ so sánh - Gabor filter, Data1



Hình 28. Biểu đồ so sánh - Chia lưới, Data1

Kết quả thí nghiệm các phương pháp sử dụng ảnh gradient thay cho ảnh lọc Gabor.

Tập kiểm tra	Độ chính xác của phương pháp (%)			
	Gabor filter	Chia lưới	Bag of words	Histogram
Data1	95.12	97.16	93.75	93.06
Data2	96.76	97.79	96.55	96.37
Data3	94.84	95.13	95.88	95.41

Bảng 4. Độ chính xác của các phương pháp khi sử dụng ảnh gradient

Tập kiểm tra	Độ chênh lệch độ chính xác (%)			
	Gabor filter	Chia lưới	Bag of words	Histogram
Data1	-0.7	-0.44	+0.45	+0.26
Data2	+0.39	+0.33	+0.21	+0.32
Data3	+0.95	-0.82	+0.41	+1.23

Bảng 5. So sánh các phương pháp với phương pháp Gradient

5. NHẬN XÉT ĐÁNH GIÁ

5.1 Ưu điểm của các phương pháp

Phương pháp Gabor filter trích xuất đặc trưng mean và variance trên toàn bộ ảnh sau khi lọc bằng bộ lọc Gabor ở 24 kênh. Điều này khiến cho độ chính xác không bị ảnh hưởng bởi khoảng cách xa gần của vật thể. Nó chỉ quan tâm đến kết cấu của ảnh tương ứng với số lượng người. Số lượng người ở 2 khung hình giống nhau thì đặc trưng trích ra gần giống nhau. Do đó, điều cần thiết là phải huấn luyện mô hình học máy có thể đảm bảo tất cả các trường hợp số lượng người có thể xảy ra. Khi điều đó được đảm bảo thì độ chính xác của phương pháp khá cao. Thời gian trích xuất đặc trưng và huấn luyện nhanh.

Phương pháp chia lưới áp dụng mặt nạ để loại bỏ những phần nằm ngoài vùng quan sát, áp dụng lưới để lấy các đặc trưng mang tính chất cục bộ địa phương. Do đó, sự thay đổi mức sáng từ vùng này sang vùng khác trên ảnh sẽ tương đương sự thay đổi thông tin mean và variance từ ô này sang ô khác trên lưới. Điều đó khiến phương pháp nhạy hơn khi ước lượng số người, độ chính xác được cải thiện hơn. Thời gian trích xuất đặc trưng và huấn luyện nhanh.

Phương pháp hoán vị thừa kế phương pháp chia lưới nhưng hóa vị các ô lưới tổng hợp làm tăng số lượng tập huấn luyện. Sự hoán vị các ô lưới giúp phương pháp này tổng quát hơn phương pháp chia lưới. Nếu hoán vị được tất cả các trường hợp ($N!$ trường hợp cho lưới N ô) thì đám đông người xuất hiện ở bất kì vị trí ô nào trong lưới thì phương pháp vẫn ước lượng chính xác được.

Phương pháp Bag of words và Histogram mang tính chất tổng quát hơn so với phương pháp Gabor filter. Cả 2 phương pháp này đều có tiềm năng cải tiến tăng thêm tính tổng quát và độ chính xác. Tính tổng quát cao nhất được hướng tới là khả năng phương pháp ước lượng có thể ước lượng chính xác khi camera đặt ở góc nhìn hoàn toàn mới mà không cần phải huấn luyện lại tập dữ liệu đã có trước đó.

Phương pháp gradient cải thiện một phần độ chính xác cho phương pháp Bag of words và phương pháp Histogram. Ảnh gradient làm nổi rõ những đường nét sau khi lọc với Gabor, tạo sự chênh lệch các giá trị trong ảnh đầu ra khiến việc phân cụm và chia bin được chính xác hơn. Các cặp giá trị (mean, variance) gần giống nhau sẽ được dễ dàng gom chung một cụm hoặc một bin hơn.

5.2 Nhược điểm của các phương pháp

Phương pháp Gabor filter và chia lưới chỉ hoạt động tốt khi đã được huấn luyện hầu hết số lượng người có thể xảy ra. Tính tổng quát chưa cao.

Phương pháp hoán vị, Bag of words và Histogram đều tốn nhiều thời gian dành trích xuất đặc trưng và huấn luyện. Đặc biệt, phương pháp hoán vị nếu chia lưới nhiều ô và hoán vị đủ tất cả các trường hợp thì thời gian để train dữ liệu là rất lớn. Phương pháp Bag of words hao tốn thời gian nhiều nhất là giai đoạn phân cụm K-means vì

lượng điểm rất lớn và thực thi 24 lần tương ứng 24 kênh. Phương pháp Histogram cũng vậy, trong hiện thực cần sử dụng Quick sort để sắp xếp thứ tự tăng dần mean để đáp ứng việc chia bin là chuẩn xác, thời gian để thực thi cũng hao tốn rất nhiều ở bước này.

6. KẾT LUẬN

6.1 Kết luận

Đề tài nghiên cứu cũng đã chỉ ra được phương pháp giúp cải thiện độ chính xác tuy độ tăng là chưa nhiều. Bên cạnh đó, mở ra tiềm năng cho các hướng phát triển để cải thiện tính tổng quát. Để xây dựng được một hệ thống ước lượng số lượng đám đông đạt tính tổng quát cao và độ chính xác chấp nhận được là rất khó khăn. Điều đó đặt ra vấn đề cho các hướng cải tiến trong tương lai.

Xét về độ chính xác, phương pháp chia lưới đạt độ chính xác cao nhất trong hầu hết các thí nghiệm. Các kết quả thí nghiệm và phân tích ưu nhược điểm cho thấy nhiều hướng đi để cải thiện phương pháp.

6.2 Kiến nghị

Phương pháp Bag of words và Histogram sử dụng lưới rất nhiều ô để đảm bảo tính thống kê. Yếu tố luật xa gần của hình ảnh so với khoảng cách camera ít nhiều ảnh hưởng đến tính tổng quát và độ chính xác của 2 phương pháp này. Độ chính xác và tính tổng quát có thể được cải thiện nếu giải quyết được 2 vấn đề như sau:

1. Yếu tố xa gần của hình ảnh
2. Số lượng người đông/thưa và phân bố bất kì vị trí nào trên ảnh đều trích xuất ra cùng một vector đặc trưng (vector đặc trưng này rất gần nhau nếu cùng nhãn và xa nhau nếu khác nhãn)

TÀI LIỆU THAM KHẢO

1. *Crowd analysis: a survey*. **Beibei, Zhan; Dorothy, Moneksso; Paolo, Remagnino; Sergio, Velastin; Li-Qun, Xu**. 2008. Machine Vision and Applications.
2. **Rafael, Gonzalez C.; Richard, Woods E.**. *Digital Image Processing*.
3. [Online] http://en.wikipedia.org/wiki/K-means_clustering.
4. *Fast Crowd Segmentation Using Shape Indexing*. **L., Dong; V., Parameswaran; V., Ramesh; I., Zoghلامي**. Rio de Janeiro, Brazil : s.n., 2007.
5. *VIDEO ANALYSIS USING CORNER MOTION STATISTICS*. **A., Albiol; M., Silla J.; A., Albiol; J., Mossi M.**. 2009. Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. pp. 31-38.
6. *A Method for Counting Moving People in Video Surveillance Videos*. **D., Conte; P., Foggia; G., Percannella; F., Tufano; M., Vento**. 2010. EURASIP Journal on Advances in Signal Processing.
7. *People Number Estimation in the Crowded Scenes Using Texture Analysis Based on Gabor Filter*. **Q., WEN; C., JIA; Y., YU; G., CHEN; Z., YU; C., ZHOU**. 2011. Journal of Computational Information Systems 7. pp. 3754-3763.
8. **Javier, Movellan R**. *Tutorial on Gabor Filters*.
9. [Online] <http://www.cvg.rdg.ac.uk/PETS2009/index.html>.
10. [Online] <http://www.esat.kuleuven.be/sista/lssvmlab/>.