

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



**ĐỒ ÁN 2**

# **TÌM HIỂU HỆ THỐNG NHẬN DẠNG ĐỐI TƯỢNG**

*Người hướng dẫn:* Ths. **VÕ HOÀNG ANH**  
*Người thực hiện:* **Nguyễn Thị Hồng - 51303396**  
**Lớp : 15030303**  
**Khoá :17**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2016**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



## ĐỒ ÁN 2

# TÌM HIỂU HỆ THỐNG NHẬN DẠNG ĐỐI TƯỢNG

*Người hướng dẫn:* Ths. **VÕ HOÀNG ANH**  
*Người thực hiện:* **Nguyễn Thị Hồng - 51303396**  
**Lớp : 15030303**  
**Khoá :17**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2016**

## **LỜI CẢM ƠN**

Em xin chân thành cảm ơn cô Võ Hoàng Anh đã tận tình hướng dẫn em hoàn thành đồ án này. Đồ án được hoàn thành nhưng kết quả không được như mong đợi do việc tìm hiểu và kiến thức của em còn hạn chế. Đồ án em làm còn nhiều sai sót, mong thầy cô bỏ qua. Em xin chân thành cảm ơn.

## **ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi và được sự hướng dẫn của Cô Võ Hoàng Anh;. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Thị Hồng*

## PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

### Phần xác nhận của GV hướng dẫn

---

---

---

---

---

---

---

Tp. Hồ Chí Minh, ngày      tháng      năm  
(kí và ghi họ tên)

### Phần đánh giá của GV chấm bài

---

---

---

---

---

---

---

Tp. Hồ Chí Minh, ngày      tháng      năm  
(kí và ghi họ tên)

## TÓM TẮT

Ngày nay, với việc phát triển không ngừng trong ngành công nghệ thông tin, nhiều ứng dụng mới được tìm hiểu và áp dụng trong đời sống. Một trong những bài toán đang được nhiều sự quan tâm hiện nay là bài toán phân loại quần áo. Việc áp dụng bài toán phân loại quần áo trong thực tế đem lại nhiều ứng dụng quan trọng. Nó không những đáp ứng được nhu cầu của con người mà còn giúp giảm nguồn nhân lực lao động. Giúp tiết kiệm được chi phí trong lao động. Vì vậy chúng em chọn đề tài này để tìm hiểu với mong muốn hiểu được cách hoạt động cũng như áp dụng nó trong đời sống hiện nay.

## MỤC LỤC

LỜI CẢM ƠN .....	i
PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN .....	iii
TÓM TẮT .....	iv
MỤC LỤC.....	1
DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ .....	3
CHƯƠNG 1 – MỞ ĐẦU .....	6
1.1 Giới thiệu .....	6
1.2 Ý nghĩa khoa học và thực tiễn .....	6
1.3 Bài toán nhận dạng đối tượng và những khó khăn .....	7
1.3.1 Các loại quần áo.....	7
1.3.2 Biểu diễn bài toán nhận dạng đối tượng trên ảnh tĩnh.....	7
1.4 Tổng quát cấu trúc hệ thống phân loại quần áo .....	7
1.5 Mục tiêu và giới hạn phạm vi đề tài.....	8
1.6 Bố cục của đồ án .....	8
CHƯƠNG 2 – CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN .....	9
CHƯƠNG 3: TIỀN XỬ LÝ DỮ LIỆU.....	10
3.1 Giới thiệu về tập dữ liệu.....	10
3.2 Xây dựng tập dữ liệu.....	10
CHƯƠNG 4 – CƠ SỞ LÝ THUYẾT .....	13
4.1 Rút trích đặc trưng bằng thuật toán Speeded Up Robust Feature.....	13
4.1.1 Ảnh tích hợp.....	13
4.1.2 Phát hiện nhanh Hessian .....	14
4.2 Phân lớp ảnh – Thuật toán K-mean .....	20
4.3 Bag of word trong phân lớp ảnh .....	22
4.3.1 Định nghĩa bag of word .....	22
4.3.2 Rút trích đặc trưng và miêu tả đặc trưng .....	23

4.3.3 Lượng tử hóa vector .....	24
4.3.4 Biểu diễn bag of word .....	25
4.4 Phương pháp máy hỗ trợ vector để phân loại .....	26
CHƯƠNG 5: MÔ TẢ BÀI TOÁN PHÂN LOẠI QUẦN ÁO .....	31
5.1 Mô hình bài toán .....	31
5.2 Mô tả bài toán phân loại quần áo .....	31
5.3 Phân lớp .....	32
5.4 Demo ứng dụng .....	33
CHƯƠNG 6: ĐÁNH GIÁ VÀ KẾT QUẢ THỰC NGHIỆM .....	35
6.1 Cách đánh giá .....	35
6.2 Kết quả thực nghiệm .....	36
CHƯƠNG 7: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....	37
7.1 Kết luận .....	37
7.2 Hướng phát triển .....	37



## DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

### DANH MỤC HÌNH

Hình 1: Ảnh gốc 1 của tập dữ liệu .....	11
Hình 2: Nhân belt được tạo từ ảnh gốc(hình 1) .....	11
Hình 3: Nhân pants được tạo từ ảnh gốc (hình 1). ....	11
Hình 4: Nhân shoes được tạo từ ảnh gốc(hình 1). ....	11
Hình 5: Ảnh gốc 2 của tập dữ liệu .....	12
Hình 6: Nhân bag được tạo từ ảnh gốc (hình 5).....	12
Hình 7: Nhân dress được tạo từ tập dữ liệu (hình 5).....	12
Hình 8: Nhân t_shirt được tạo từ tập dữ liệu (hình 5).....	12
Hình 9: Hình I có ma trận đầu vào và kết quả sau khi tính ảnh tích hợp.....	14
Hình 10: Biểu đồ tỉ lệ phát hiện số lượng điểm quan tâm được phát hiện trên mỗi octave phân rã nhanh chóng. ....	15
Hình 11: Những điểm quan tâm được phát hiện trên một cánh đồng hoa hướng dương. ....	15
Hình 12: Phép lọc Haar wavelet để tính toán đặc trưng ở x (bên trái) và y hướng (bên phải). Vùng đen có trọng số -1 và vùng trắng có trọng số +1.....	16
Hình 13: Gán hướng cho điểm quan tâm. ....	17
Hình 14: Chi tiết của ảnh Graffiti thể hiện kích thước của cửa sổ bộ mô tả hướng ở các tỉ lệ khác nhau. ....	17
Hình 15: Xây dựng bộ mô tả và hướng của điểm quan tâm. ....	18
Hình 16: Các mục mô tả của miền con đại diện cho tính chất của mẫu cường độ cơ bản. ....	19
Hình 17: Nếu độ tương phản giữa hai điểm quan tâm khác nhau, ứng viên sẽ không được xem là so khớp có giá trị. ....	20
Hình 18: Gom cụm bằng giải thuật K-means .....	21
Hình 19: Bag of word của một ảnh .....	22

Hình 20: Đặc trưng của một ảnh không phụ thuộc lẫn nhau .....	22
Hình 21: Biểu đồ thể hiện tần số xuất hiện các đặc trưng. ....	23
Hình 22: Bộ từ điển đối tượng. ....	23
Hình 23: Rút trích đặc trưng ảnh.....	23
Hình 24: Quá trình gom cụm, lượng tử hóa vector. ....	24
Hình 25: Từ điển những từ được mã hóa. ....	24
Hình 26: Những điểm ảnh của mã hóa vector. ....	25
Hình 27: Cách tạo ra từ điển bằng cách lượng tử hóa vector.....	25
Hình 28: Tần số xuất hiện các đặc trưng.....	26
Hình 29: Thuật toán SVM dùng để phân lớp 2 tập $x+$ và $x-$ .....	26
Hình 30: Tập dữ liệu được phân chia tuyến tính.....	27
Hình 31: Tập dữ liệu phân chia tuyến tính nhưng có nhiễu. ....	28
Hình 32: Tập dữ liệu không phân chia tuyến tính được. ....	29
Hình 33: Biểu diễn tập dữ liệu trên không gian 2 chiều. ....	29
Hình 34: Mô hình phân loại quần áo.....	31
Hình 35: Mô hình quá trình phân lớp.....	33
Hình 36: Giao diện demo ứng dụng. ....	33
Hình 37: Chọn ảnh cần phân loại. ....	33
Hình 38: Nhãn cần phân loại được nhận dạng đúng. ....	34
Hình 39: Nhãn cần phân loại nhận dạng bị sai từ “Suit” thành “T-shirt”.....	34
Hình 40: Mô tả cách chia tập huấn luyện và tập kiểm tra.....	35

## **DANH MỤC BẢNG**

Bảng 1: Các loại quần áo có trong tập dữ liệu phân loại .....	7
Bảng 2: Các công trình nghiên cứu liên quan .....	9
Bảng 3: Các nhãn có trong bài công trình nghiên cứu Clothing Co-Parsing.....	10
Bảng 4: Các nhãn ở trong tập dữ liệu lấy từ công trình nghiên cứu trên.....	11
Bảng 5: Kết quả thực nghiệm sau 10 lần chạy.....	36

# CHƯƠNG 1 – MỞ ĐẦU

## 1.1 Giới thiệu

Quần áo là đối tượng được con người sử dụng dùng để mặc, mang trên người. Ban đầu vai trò của quần áo là che thân và bảo vệ cơ thể còn bây giờ nó đã vươn lên thành thời trang thể hiện cái đẹp, nét văn hóa và phong cách của người mặc. Ngày nay trang phục cũng trở lên đa dạng, phong phú hơn. Cho nên trang phục có nhiều loại như là áo sơ mi, quần tây, áo vest, áo dài, đầm, chân váy, áo khoác, áo thun, quần đùi, áo len, đồ thể thao và còn thêm một số phụ kiện kèm theo như là vòng tay, khuyên tai, nhẫn, kính, găng tay, giày, dép, ví, đồng hồ, túi xách .... Do trang phục đa dạng như vậy nên công việc phân loại chúng rất tốn chi phí, mất thời gian và sức lực. Khi đó câu hỏi được đặt ra là làm thế nào để phân loại chúng vừa nhanh mà lại tiết kiệm? Và bài toán phân loại quần áo đã ra đời. Bài toán thuộc lĩnh vực thị giác máy tính đã và đang ngày càng thu hút được nhiều sự quan tâm của các nhà nghiên cứu. Các công trình nghiên cứu liên quan như là DeepFashion - Clothing recognition and retrieval (Ziwei Liu, Ping Luo...) hay là Parsing Clothing in Fashion Photographs (Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg).

Cũng tương tự như vấn đề của các công trình trên trong bài này là bài toán phân loại quần áo với đầu vào là một hình ảnh chứa thông tin là một người và đầu ra của nó là người đó đang mặc trên người những loại quần áo nào.

Các hướng tiếp cận bài toán là các giải thuật thuộc lĩnh vực thị giác máy tính như là mô hình **Bag of word (BoW)**, phương pháp gom cụm **k-nearest neighbors (KNN)**, phương pháp phân lớp **support vector machine (SVM)**.

## 1.2 Ý nghĩa khoa học và thực tiễn

Do trang phục ngày càng phong phú đa dạng và nhu cầu con người ngày càng cao. Để đáp ứng nhu cầu cũng như sở thích của con người, ta phải đặt ra những câu hỏi như là họ cần gì và muốn gì để rồi đáp ứng những điều đó ra sao. Thì bài này chính là bước tiền xử lý cho việc giải đáp những câu hỏi đó.

Sự phát triển của bài toán là cơ sở phát triển nhiều ứng dụng hiện đại ngày nay đặc biệt trong công nghệ chế tạo robot – một trong những lĩnh vực mang nhiều hứa hẹn trong tương lai.

▪ ***Bài toán ứng dụng trong các lĩnh vực như:***

Phân loại quần áo - cho biết một người họ mang đồ gì và họ mặc trang phục như thế nào.

Phân tích quần áo - là những tính chất, đặc trưng mạnh, thích ứng với những đối tượng, thời điểm cụ thể.

Truy hồi thông tin quần áo - Dựa vào những đặc trưng, tính chất trong một đối tượng quần áo ta tìm những đối tượng tương tự.

Và rộng hơn sẽ được ứng dụng vào lĩnh vực công nghiệp sản xuất, chế tạo sản phẩm may mặc.

### **1.3 Bài toán nhận dạng đối tượng và những khó khăn**

#### ***1.3.1 Các loại quần áo***

Bag	Belt	Blazer	Blouse	Coat
Dress	Hat	Jacket	Jeans	Pants
Shoes	Shorts	Suit	Sweater	T-shirt

Bảng 1: Các loại quần áo có trong tập dữ liệu phân loại

#### ***1.3.2 Biểu diễn bài toán nhận dạng đối tượng trên ảnh tĩnh***

Đầu vào: một hình ảnh I

Đầu ra: hình ảnh đó được phân loại thành một trong nhóm 5 loại trên.

### **1.4 Tổng quát cấu trúc hệ thống phân loại quần áo**

Bao gồm 4 bước sau:

Bước 1: Tiền xử lý dữ liệu là tạo tập dữ liệu từ công trình nghiên cứu **Clothing Co-Parsing (CCP)**.

Bước 2: Tiền xử lý ảnh là quá trình tìm kiếm và tìm thấy đối tượng phân loại trong một ảnh.

Bước 3: Rút trích và lựa chọn đặc trưng của loại.

Ta xác định ảnh đó thuộc loại nào, sau đó rút trích nhưng đặc trưng mạnh, quan trọng. Trong đề tài này sẽ sử dụng thuật toán SURF để tìm điểm đặc trưng.

Bước 4: Gom những đối tượng có liên quan về cùng một cụm bằng thuật toán k-means và phân lớp dựa trên giải thuật SVM.

### **1.5 Mục tiêu và giới hạn phạm vi đề tài**

- Tạo được tập dữ liệu riêng và có độ chính xác cao.
- Thực hiện được phương pháp rút trích đặc trưng SURF một cách hiệu quả cho bài toán.
- Đồ án tập trung tìm hiểu và xây dựng hệ thống phân loại quần áo với tập ảnh tích hợp để giúp cho việc nhận dạng, phân loại dễ dàng hiệu quả hơn.
- Phạm vi đồ án giới hạn tìm hiểu trên đối tượng ảnh là vùng quần áo.

### **1.6 Bố cục của đồ án**

- Chương 1: Mở đầu, giới thiệu bài toán phân loại quần áo
- Chương 2: Các công trình nghiên cứu liên quan đến bài toán phân loại quần áo.
- Chương 3: Tiền xử lý dữ liệu.
- Chương 4: Cơ sở lý thuyết.
- Chương 5: Xây dựng mô hình phân loại quần áo.

## CHƯƠNG 2 – CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

STT	Tên bài báo	Tác giả – năm	Phương pháp	Tập dữ liệu	Tỉ lệ đúng
1	Parsing Clothing in Fashion Photographs <a href="#">[1]</a>	Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg (2012)	Tạo supperpixels, từ đó ước lượng tư thế con người và dự đoán nhãn trong một tập L	Novel dataset chứa 158,235 hình ảnh thời trang	80.8% pixel accuracy
2	A High Performance CRF Model for Clothes Parsing <a href="#">[2]</a>	Edgar Simo-Serra, Sanja Fidler,Francesc Moreno - Noguera, Raquel Urtasun (2014)	Sử dụng mô hình Conditional Random Field (CRF) để lựa chọn các đặc trưng khác nhau và quan trọng từ tập dữ liệu và dùng các đặc trưng này để phân loại	Novel dataset chứa 158,235 hình ảnh thời trang	84.88 pixel accuracy
3	Clothing Co- Parsing by Joint Image Segmentation and Labeling <a href="#">[3]</a>	Wei Yang, Ping Luo, Liang Lin (16/3/2016)	Sử dụng SVM để trích xuất các hình ảnh đồng vùng và xây dựng mô hình nhãn dựa vào vùng phân đoạn như đỉnh , bối cảnh, hình dạng quần áo.	2098 hình ảnh đường phố có độ phân giải cao	90.29% segmentation accuracy

Bảng 2: Các công trình nghiên cứu liên quan

## CHƯƠNG 3: TIỀN XỬ LÝ DỮ LIỆU

### 3.1 Giới thiệu về tập dữ liệu

Tập dữ liệu trong bài này được lấy từ công trình nghiên cứu **Clothing Co-Parsing (CCP)** [3] là một công trình nghiên cứu về phân tích những đặc trưng tính chất của quần áo, từ đó sẽ chia ra những phân đoạn vùng ảnh. Kết quả của tập dữ liệu này gồm 2098 hình ảnh đường phố có độ phân giải cao - hình ảnh có tính đa dạng về phong cách, phụ kiện, quần áo, và các tư thế tạo dáng khác nhau.

Trong đó 1004 hình đầu được sử dụng làm pixel-level annotations và 1094 hình còn lại sử dụng làm image-level annotations.

Tập dữ liệu của bài được xử lý từ 1004 tấm hình pixel-level annotations bằng cách tách các vùng của cùng một loại với nhau, sau đó phân thành những đối tượng cùng loại và tập các tập train và test được sử dụng từ tập này.

### 3.2 Xây dựng tập dữ liệu

Tập dữ liệu lấy từ công trình nghiên cứu có 53 nhãn:

Accessories	Bag	Belt	Blazer	Blouse	Bodysuit
Boots	Bra	Bracelet	Earring	Flats	Glasses
Gloves	Hair	Hat	Heels	Hoodie	Intimate
Jacket	Jeans	Jumper	Leggings	Loafers	Necklace
Panties	Pants	Pumps	Purse	Ring	Romper
Sandals	Scarf	Shirt	Shoes	Shorts	Skin
Skirt	Sneakers	Socks	Stockings	Suit	Sunglasses
Sweater	Sweatshirt	Swimwear	t-shirt	Tie	Tights
Top	Vest	Wallet	Watch	Wedges	

Bảng 3: Các nhãn có trong bài công trình nghiên cứu Clothing Co-Parsing.



Nhưng để phù hợp với bài toán phân loại quần áo này, cho nên sau khi lọc và xử lý những nhãn không phù hợp với nhiều nhãn có số lượng dữ liệu quá ít nên tập dữ liệu cuối cùng còn 15 nhãn như sau:

Bag (92 hình)	Belt (97 hình)	Blazer (83 hình)	Blouse (89 hình)	Coat (126 hình)
Dress (101 hình)	Hat (88 hình)	Jacket (59 hình)	Jeans (95 hình)	Pants (125 hình)
Shoes (86 hình)	Shorts (57 hình)	Suit (78 hình)	Sweater (64 hình)	T-shirt (57 hình)

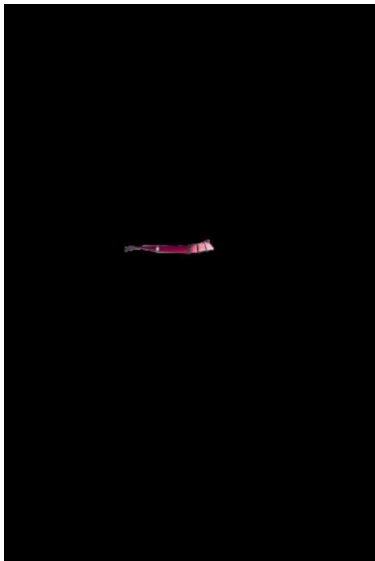
Bảng 4: Các nhãn ở trong tập dữ liệu lấy từ công trình nghiên cứu trên.



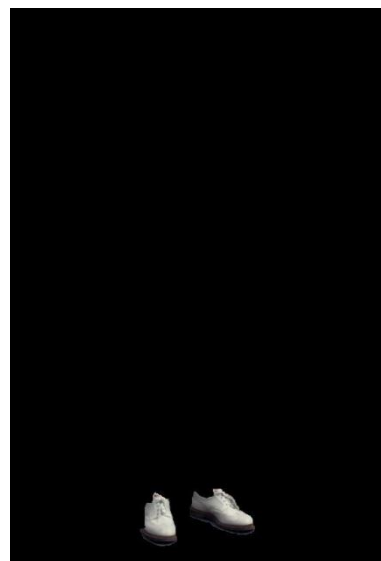
Hình 1: Ảnh gốc 1 của tập dữ liệu



Hình 3: Nhãn pants được tạo từ ảnh gốc (hình 1).



Hình 2: Nhãn belt được tạo từ ảnh gốc(hình 1)



Hình 4: Nhãn shoes được tạo từ ảnh gốc(hình 1).



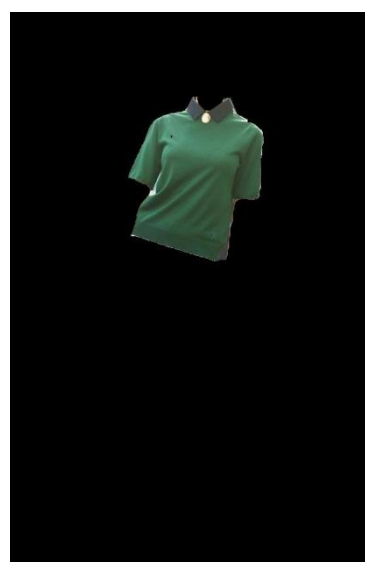
Hình 5: Ảnh gốc 2 của tập dữ liệu



Hình 7: Nhãn dress được tạo từ tập dữ liệu (hình 5)



Hình 6: Nhãn bag được tạo từ ảnh gốc (hình 5)



Hình 8: Nhãn t\_shirt được tạo từ tập dữ liệu (hình 5)

Tập dữ liệu được tạo từ 1004 ảnh gốc thành các bộ phận của quần áo, trang phục sau đó chúng được gán thành các nhãn phân loại dựa theo 15 nhãn trên.

## CHƯƠNG 4 – CƠ SỞ LÝ THUYẾT

### 4.1 Rút trích đặc trưng bằng thuật toán Speeded Up Robust Feature

Thuật toán Speeded Up Robust Features (SURF) được giới thiệu năm 2006 bởi nhóm nghiên cứu của Herbert Bay [4]. Ý tưởng của thuật toán này được bắt nguồn và phát triển từ thuật toán SIFT nên tương đối giống với thuật toán SIFT.

SURF cũng dùng không gian tỉ lệ để tìm kiếm điểm đặc trưng và các đặc trưng được mô tả bởi các vector có đi kèm thêm hướng. Thuật toán này có hai phần chính là “phát hiện” điểm đặc trưng, và “mô tả” điểm đặc trưng.

Bộ phát hiện đặc trưng của SURF: quá trình phát hiện điểm đặc trưng dựa trên không gian tỉ lệ, và sử dụng kết quả xấp xỉ trên ma trận Hessian và ảnh tích hợp để đơn giản hơn trong việc tính toán). Các điểm đặc trưng là các điểm cực đại của hàm Hessian. Người ta gọi bước này là phát hiện nhanh Hessian (*Fast Hessian Detector*).

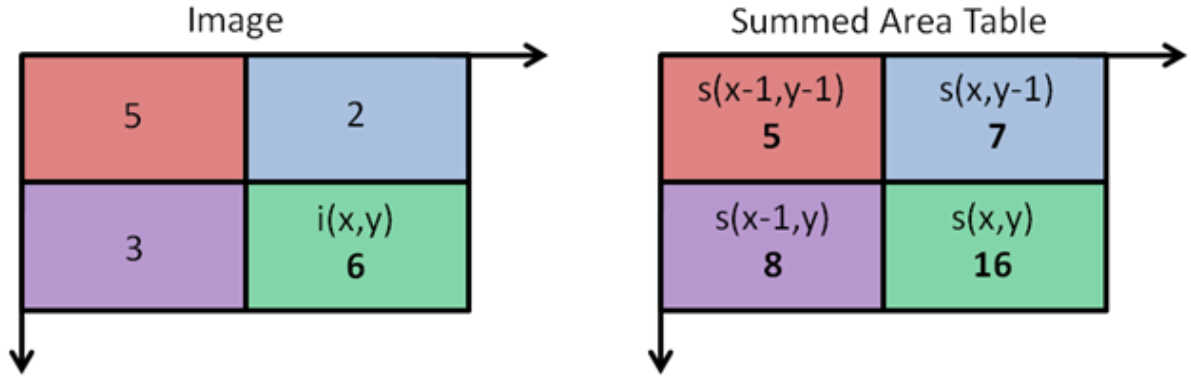
Mô tả: mỗi điểm đặc trưng được thêm một hướng dùng để nhận dạng khi ảnh bị xoay, và được mô tả bằng một vector gồm 64 chiều chứa thông tin biến thiên trên ảnh. Cả hai bước trên đều sử dụng ảnh tích hợp để tăng tốc độ tính toán.

#### 4.1.1 Ảnh tích hợp

Ảnh tích hợp, được đưa ra bởi Viola và Jones là một mảng 2 chiều có kích thước bằng kích thước ảnh cần rút trích đặc trưng. Mỗi phần tử của mảng được tính bằng tổng của điểm ảnh phía trên (dòng -1) và bên trái (cột -1) của nó. Bắt đầu ở vị bên trên, phía bên trái đến vị trí bên dưới, phía bên phải của ảnh, việc tính toán này được thực hiện bằng cách dựa trên phép cộng các số nguyên vì vậy tốc độ thực hiện rất nhanh

$$I_{\Sigma}(X) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

Ví dụ về tính ảnh tích hợp: Hình I có ma trận như sau



Hình 9: Hình I có ma trận đầu vào và kết quả sau khi tính ảnh tích hợp.

#### 4.1.2 Phát hiện nhanh Hessian

Việc phát hiện đặc trưng của thuật toán SURF dựa vào ma trận Hessian, bởi vì hiệu suất tốt của nó về mặt thời gian cũng như độ chính xác. Thay vì sử dụng một biện pháp khác để lựa chọn vị trí và tỉ lệ (như đã được thực hiện trong phát hiện Hessian-Laplace [5]). Mặt khác, mô tả này mô tả một phân bố các đặc trưng Haar-wavelet trong các láng giềng của điểm quan tâm. Hơn nữa, chỉ sử dụng kích thước 64 để giảm thời gian tính toán các đặc trưng và so khớp, đồng thời tăng sự chắc chắn.

Với mỗi điểm  $X(x, y)$  của ảnh I, ma trận Hessian của X có tỉ lệ  $\sigma$  được định nghĩa như sau:

❖ Ma trận Hessian

$$H(X, \sigma) = \begin{bmatrix} Lxx(X, \sigma) & Lxy(X, \sigma) \\ Lxy(X, \sigma) & Lyy(X, \sigma) \end{bmatrix} \quad (3.2)$$

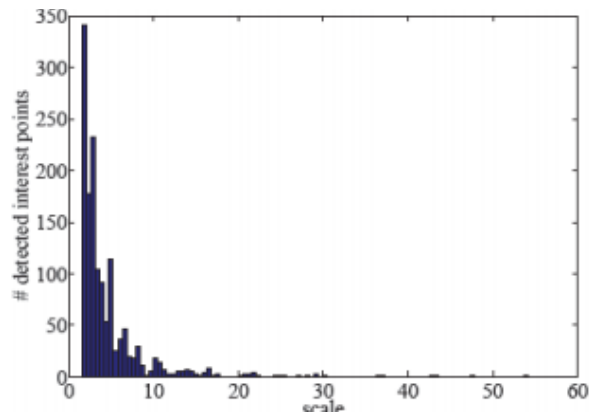
$Lxx(\mathbf{x}, \sigma)$ : tích chập đạo hàm bậc hai x của hàm Gaussian  $\frac{\partial^2}{\partial x^2} g(\sigma)$  tại điểm X của ảnh I

$Lxy(\mathbf{x}, \sigma)$ : tích chập đạo hàm bậc hai theo xy của hàm Gaussian  $\frac{\partial^2}{\partial x^2} g(\sigma)$  tại điểm X của ảnh I

$Lyy(y, \sigma)$ : tích chập đạo hàm bậc hai theo  $y$  của hàm Gaussian  $\frac{\partial^2}{\partial x^2} g(\sigma)$  tại điểm  $Y$  của ảnh  $I$

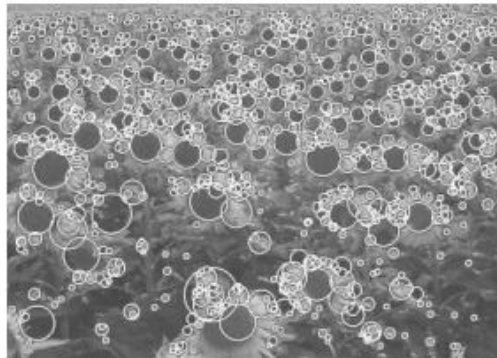
❖ Định vị điểm quan tâm

Để định vị điểm quan tâm trong ảnh và trên các tỷ lệ, một sự loại trừ không cực đại trong một láng giềng  $3 \times 3 \times 3$  được áp dụng, cực đại của định thức ma trận Hessian được nội suy trong tỷ lệ và không gian ảnh.



Hình 10: Biểu đồ tỉ lệ phát hiện số lượng điểm quan tâm được phát hiện trên mỗi octave phân rã nhanh chóng.

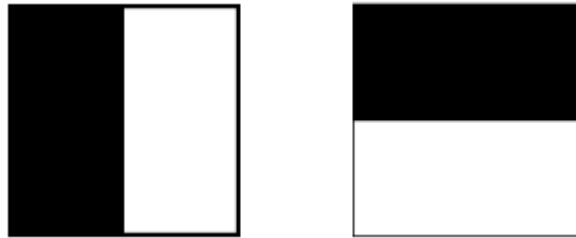
Nội suy không gian tỷ lệ là đặc biệt quan trọng trong trường hợp này, như sự khác biệt tỷ lệ giữa các lớp đầu tiên của mỗi octave là tương đối lớn. Hình trên cho thấy một ví dụ về các điểm quan tâm được phát hiện bằng cách sử dụng phát hiện Fast-Hessian.



Hình 11: Những điểm quan tâm được phát hiện trên một cánh đồng hoa hướng dương.

### ❖ Gán hướng

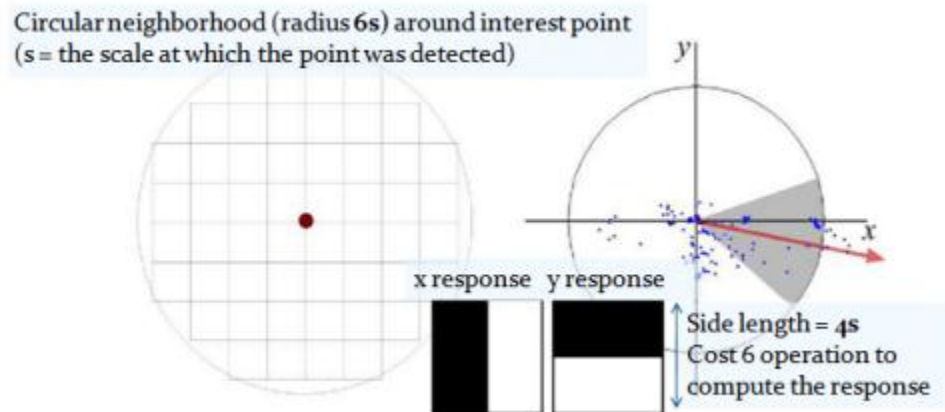
Để bắt đầu với phép xoay ảnh, ta định nghĩa một hướng lặp cho các điểm quan tâm. Với mục đích đó, đầu tiên ta tính toán đặc trưng Haar-wavelet ở x và y hướng, những đặc trưng này nằm trong vòng tròn láng giềng bán kính  $6s$  xung quanh điểm quan tâm, với  $s$  là tỷ lệ các điểm quan tâm được phát hiện. Bước lấy mẫu phụ thuộc tỷ lệ được chọn để có  $s$ . Để phù hợp với phần còn lại, kích thước các wavelet là phụ thuộc tỷ lệ, được thiết lập độ dài cạnh là  $4s$ . Vì vậy ta có thể sử dụng ảnh tích hợp cho việc lọc nhanh hơn. Các phép lọc sử dụng được chỉ ra trong hình dưới. Chỉ 6 phép toán được sử dụng để tính toán đặc trưng ở x hoặc y hướng ở bất kỳ tỷ lệ nào.



Hình 12: Phép lọc Haar wavelet để tính toán đặc trưng ở x (bên trái) và y hướng (bên phải). Vùng đen có trọng số  $-1$  và vùng trắng có trọng số  $+1$ .

Mỗi một đặc trưng wavelet được tính toán và điều chỉnh với hàm Gaussian ( $\sigma = 2.5s$ ) ở tâm điểm của điểm quan tâm, các đặc trưng được biểu diễn như vector trong không gian với độ lớn đặc trưng theo chiều ngang dọc theo hoành độ và độ lớn đặc trưng thẳng dọc theo tung độ. Hướng chủ đạo được ước tính bằng cách tính tổng tất cả các đặc trưng trong một cửa sổ định hướng trượt bao quanh một góc  $\frac{\pi}{3}$ , xem hình dưới.

Các đặc trưng ngang và dọc trong cửa sổ được tính tổng, hai đặc trưng được tổng hợp sinh ra một vector mới. Hướng vector dài nhất tương ứng với hướng điểm quan tâm. Kích thước cửa sổ trượt là một đối số được chọn cẩn thận.



Hình 13: Gán hướng cho điểm quan tâm.

❖ Bộ mô tả dựa trên tổng các đặc trưng Haar wavelet

Để tách mô tả, bước đầu tiên bao gồm việc xây dựng một khu vực hình vuông bao quanh điểm quan tâm và hướng theo hướng đã chọn trong phần trước. Kích thước cửa sổ này là  $20s$ . Ví dụ về các vùng vuông được minh họa như hình dưới.

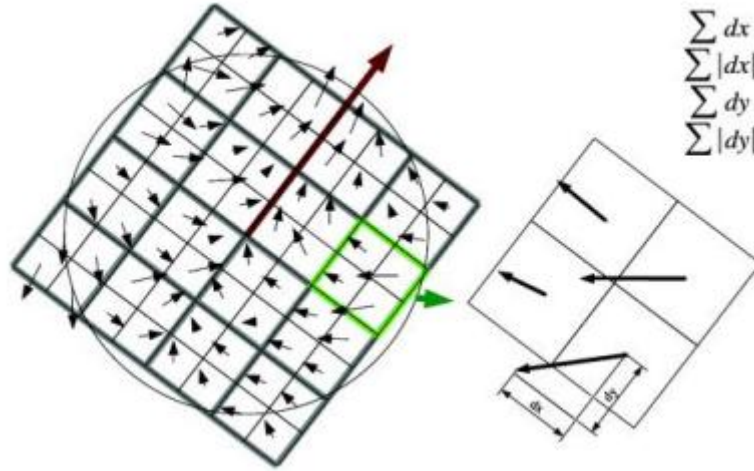


Hình 14: Chi tiết của ảnh Graffiti thể hiện kích thước của cửa sổ bộ mô tả hướng ở các tỉ lệ khác nhau.

Khu vực này thường được chia nhỏ thành các vùng con  $4 \times 4$ . Điều này giúp giữ được các thông tin không gian quan trọng. Đối với từng khu vực con, ta tính toán những đặc trưng đơn của các điểm mẫu cách quãng  $5 \times 5$ . Để đơn giản, ta gọi  $dx$  là đặc trưng Haar wavelet theo phương nằm ngang và  $dy$  là đặc trưng Haar wavelet theo



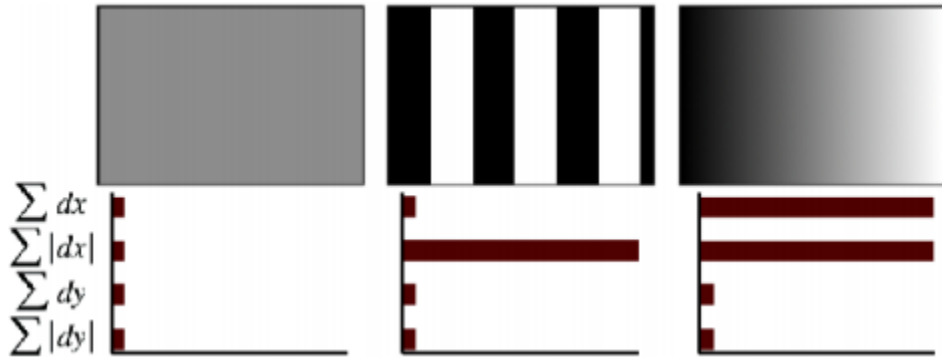
phương thẳng đứng (bộ lọc kích thước 2s). “Ngang” và “thẳng đứng” ở đây được định nghĩa theo hướng lựa chọn các điểm quan tâm. Để gia tăng độ chắc chắn đối với biến dạng hình học và các lỗi định vị, các đặc trưng  $dx$ ,  $dy$  là trọng số đầu tiên với hàm Gaussian ( $\sigma = 3.3s$ ) có tâm tại điểm quan tâm.



Hình 15: Xây dựng bộ mô tả và hướng của điểm quan tâm.

Để xây dựng bộ mô tả, một khung lưới hướng bậc hai với 4x4 vùng con hình vuông được đặt trên các điểm quan tâm (bên trái) xem hình trên. Ở mỗi hình vuông đặc trưng wavelet được tính toán. Sự chia nhỏ 2x2 của mỗi hình vuông tương ứng với các trường thực của bộ mô tả. Những cái này là tổng của  $dx$ ,  $dy$ ,  $|dx|$  và  $|dy|$ , tính toán tương đối hướng của lưới (bên phải) xem hình trên.

Sau đó các đặc trưng wavelet  $dx$ ,  $dy$  được tổng hợp qua từng khu vực con và hình thành một tập chỉ mục vector đặc trưng đầu tiên. Để mang lại thông tin về sự khác biệt của thay đổi cường độ, ta trích tổng trị tuyệt đối các đặc trưng  $|dx|$ ,  $|dy|$ . Do đó, một phân miền sẽ có một vector mô tả 4 chiều  $v$  cho cấu trúc cường độ cơ bản của phân miền  $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$ . Kết quả là một vector mô tả cho tất cả 4x4 vùng con có độ dài 64. Bất biến tương phản (một thừa số tỷ lệ) đạt được bằng cách chuyển mô tả vào trong một vector đơn vị.



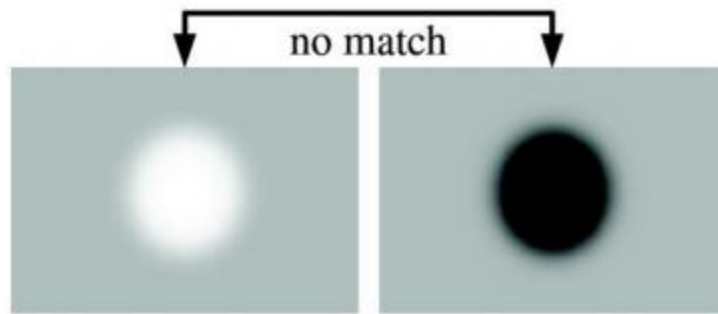
Hình 16: Các mục mô tả của miền con đại diện cho tính chất của mẫu cường độ cơ bản.

Hình trái: trong trường hợp của một miền con đồng nhất, tất cả các giá trị là tương đối thấp. Hình giữa: sự có mặt của tần số trong x hướng, giá trị của  $\sum |dx|$  là cao, nhưng tất cả các trường hợp khác vẫn thấp. Nếu cường độ tăng dần theo x hướng, cả  $\sum dx$  và  $\sum |dx|$  đều cao.

Hình trên cho thấy các đặc tính của mô tả trong ba ảnh cường độ mẫu khác nhau trong một phân miền. Ta có thể hình dung kết quả của việc kết hợp các mẫu cường độ cục bộ là một mô tả đặc biệt.

#### ❖ Lập chỉ mục và so khớp

Để lập chỉ mục nhanh chóng trong giai đoạn so khớp, dấu hiệu Laplacian (nghĩa là dấu vết của ma trận Hessian) là bao gồm cho các điểm quan tâm cơ bản. Thông thường các điểm quan tâm được tìm thấy ở các cấu trúc kiểu đốm màu (blob-type). Dấu hiệu Laplacian khác biệt các đốm màu sáng trên nền tối từ trạng thái nghịch đảo. Đặc điểm này có giá trị ở chỗ không thêm chi phí tính toán vì nó đã được tính toán trong giai đoạn phát hiện. Ở giai đoạn so khớp, ta chỉ so sánh các đặc trưng nếu nó có cùng kiểu tương phản, như ở hình dưới. Do đó, thông tin tối thiểu này cho phép so khớp nhanh hơn, mà không làm giảm hiệu suất của bộ mô tả. Lưu ý rằng đây cũng là lợi thế cho các phương pháp lập chỉ mục nâng cao hơn. Ví dụ với k-d tree, thông tin bổ sung này định nghĩa một siêu phẳng có ý nghĩa cho việc tách dữ liệu, trái với việc chọn ngẫu nhiên một phần tử hay thống kê các đặc trưng.



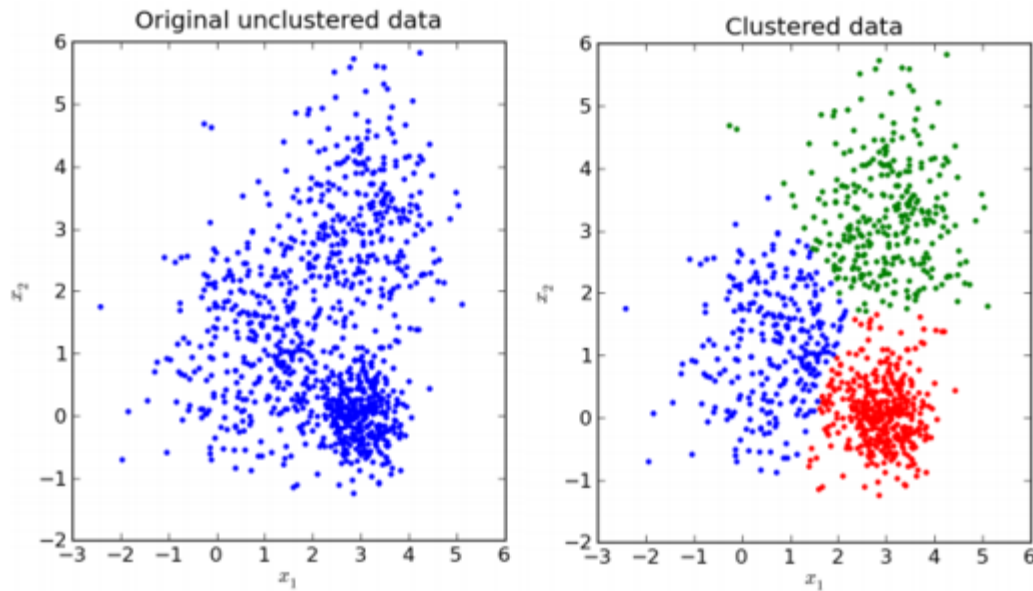
Hình 17: Nếu độ tương phản giữa hai điểm quan tâm khác nhau, ứng viên sẽ không được xem là so khớp có giá trị.

## 4.2 Phân lớp ảnh – Thuật toán K-mean

Trong thực tế chúng ta không so khớp ảnh trực tiếp với các đặc trưng của hai ảnh với nhau vì không gian và thời gian tìm kiếm lớn, mà sử dụng một phương pháp phân lớp rồi sau đó mới thực hiện so khớp ảnh. Thuật toán k-means là thuật toán gom cụm với số lần lặp xác định để tạo ra k cluster bằng cách tính trung bình khoảng cách giữa các vector đặc trưng gần nhất. Tập hợp các cluster gọi là visual vocabulary, kích thước của vocabulary phụ thuộc vào số lượng các cluster.

- Mô tả giải thuật:
  - ✓ Bước 1: Chọn ngẫu nhiên k data point để làm k điểm trọng tâm  $m_k$ , mỗi điểm trọng tâm sẽ xác định một tập con.
  - ✓ Bước 2: Lần lượt tính khoảng cách giữa các data point  $x_i$  trong tập dữ liệu đang xét với từng điểm trọng tâm  $m_k$ , điểm nào gần trọng tâm  $m_k$  nhất sẽ thuộc về tập hợp k. Khoảng cách đó có tính bằng công thức Euclide như trên hoặc Manhattan
  - ✓ Bước 3: Trong mỗi tập con ta lần lượt tính lại trọng tâm mới bằng cách lấy trung bình cộng của tất cả các data point.
  - ✓ Bước 4: Lặp lại bước 2 và 3 cho đến khi nào tọa độ của tất cả các điểm trọng tâm không còn thay đổi nữa hoặc sau một số bước đủ lớn ( ví dụ như 200 lần)

✓ Bước 5: Sau đó mỗi đặc trưng sẽ được gán vào các cluster gần nhất như hình



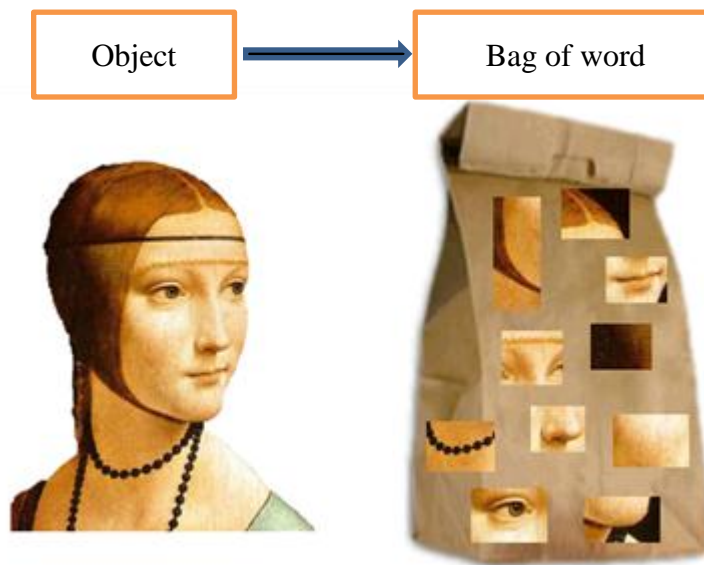
Hình 18: Gom cụm bằng giải thuật K-means

Thuật toán K-Means có ưu điểm là đơn giản, dễ hiểu và cài đặt. Tuy nhiên, một số hạn chế của K-Means là hiệu quả của thuật toán phụ thuộc vào việc chọn số nhóm K (phải xác định trước) và chi phí cho thực hiện vòng lặp tính toán khoảng cách lớn khi số cụm K và dữ liệu phân cụm lớn.

Việc tìm ra những điểm ảnh tương đối có thể chia làm 3 bước:

- ❖ Việc lựa chọn những điểm ở vị trí phân biệt trong ảnh
- ❖ Vùng xung quanh của điểm đặc trưng được xác định và mô tả bởi vector đặc trưng.
- ❖ Những vector đặc trưng được so trùng giữa nhiều tấm ảnh khác nhau.

### 4.3 Bag of word trong phân lớp ảnh



Hình 19: Bag of word của một ảnh

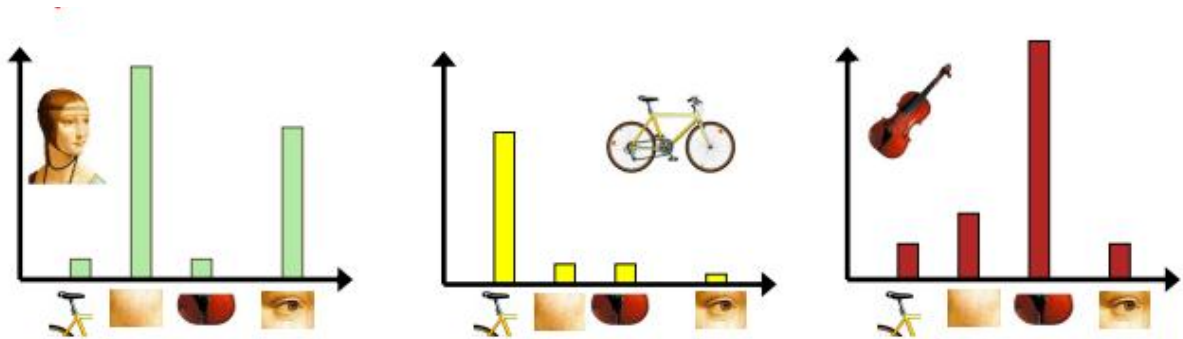
#### 4.3.1 Định nghĩa *bag of word*

- Là những đặc trưng không phụ thuộc lẫn nhau

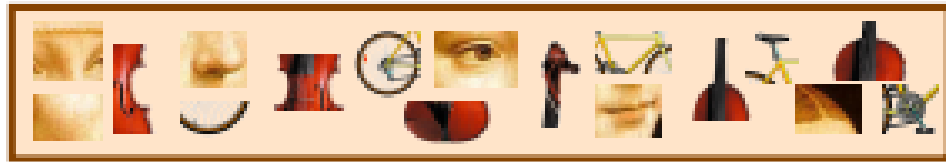


Hình 20: Đặc trưng của một ảnh không phụ thuộc lẫn nhau

- Thể hiện dưới dạng biểu đồ



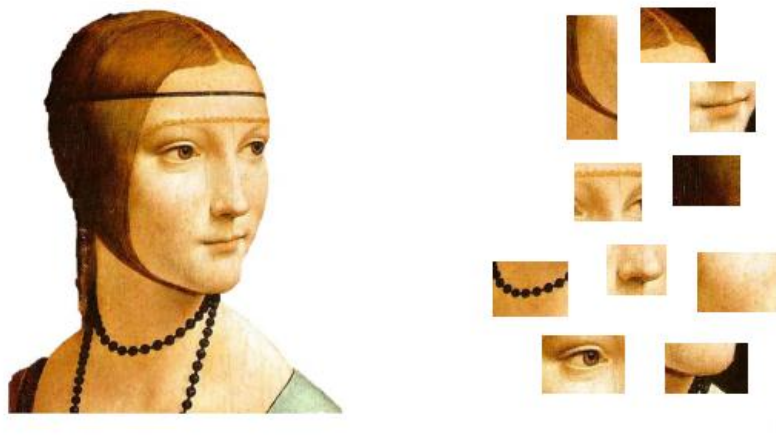
Hình 21: Biểu đồ thể hiện tần số xuất hiện các đặc trưng.



Hình 22: Bộ từ điển đối tượng.

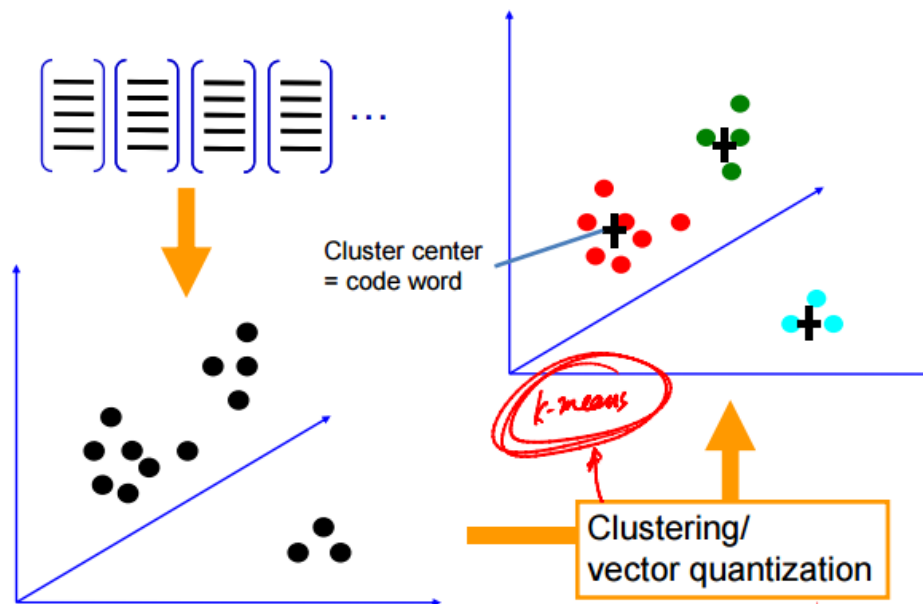
#### 4.3.2 Rút trích đặc trưng và miêu tả đặc trưng

- Rút trích đặc trưng

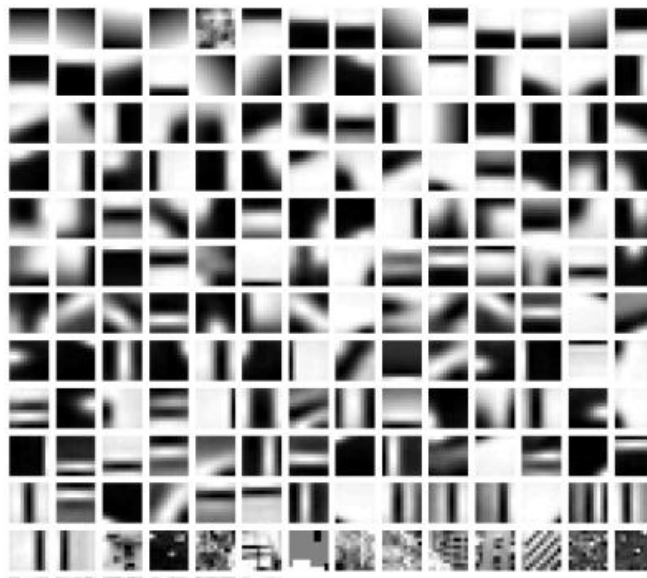


Hình 23: Rút trích đặc trưng ảnh

### 4.3.3 Lượng tử hóa vector

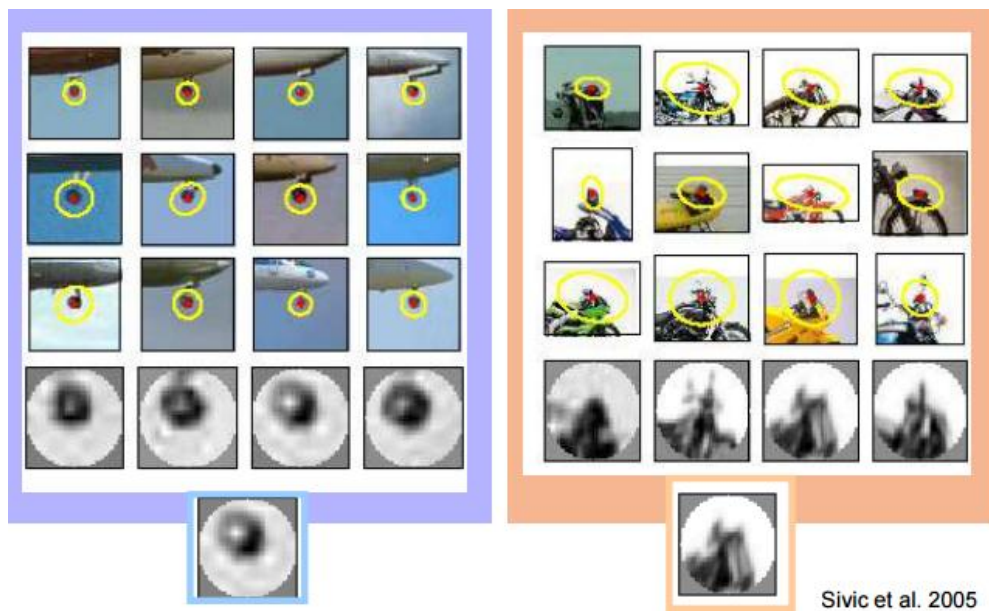


Hình 24: Quá trình gom cụm, lượng tử hóa vector.



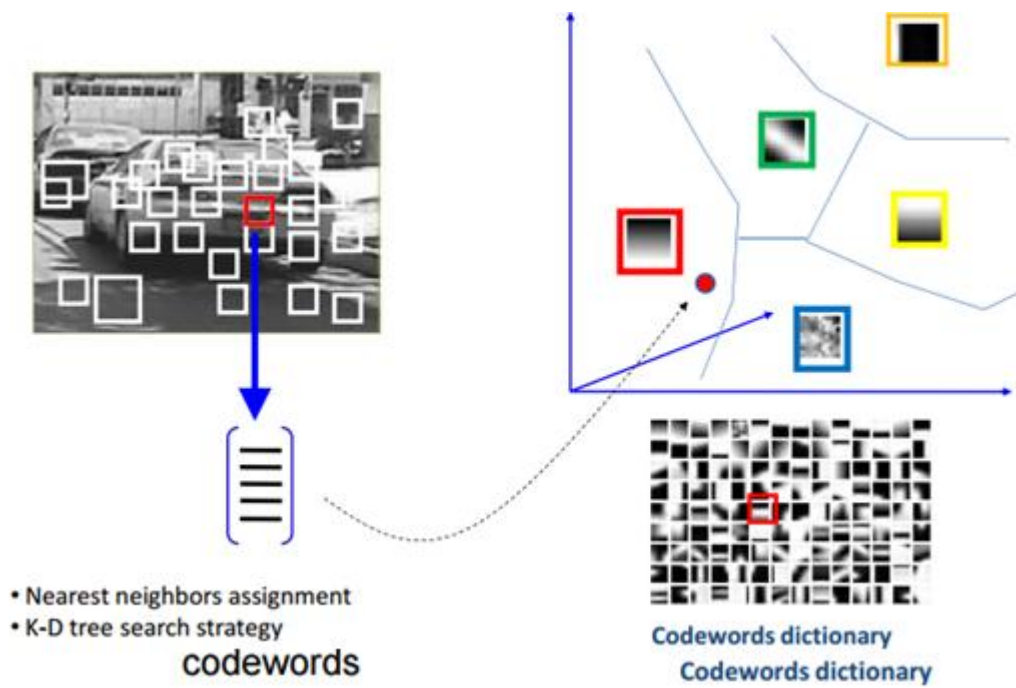
Hình 25: Từ điển những từ được mã hóa.





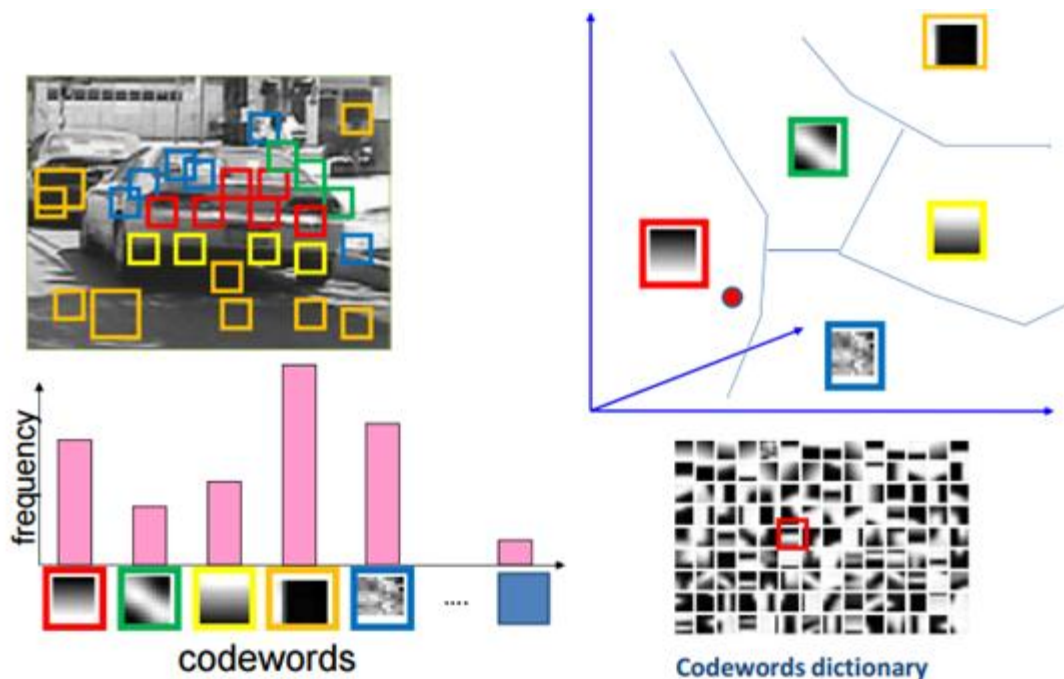
Hình 26: Những điểm ảnh của mã hóa vector.

#### 4.3.4 Biểu diễn bag of word



Hình 27: Cách tạo ra từ điển bằng cách lượng tử hóa vector.





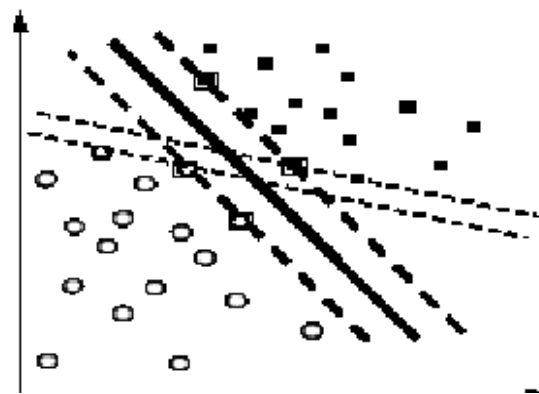
Hình 28: Tần số xuất hiện các đặc trưng.

#### 4.4 Phương pháp máy hỗ trợ vector để phân loại

##### ▪ Máy hỗ trợ vector tuyến tính

Ý tưởng của thuật toán: cho tập hợp gồm 2 loại phần tử + và - . Mỗi phần tử  $x_-$ ,  $x_+$  là vector  $n$  chiều. Thuật toán được áp dụng để tìm giải pháp ngăn cách 2 tập hợp trên.

Đường tô đậm là siêu phẳng tốt nhất và các điểm được bao bởi hình chữ nhật là những điểm gần siêu phẳng nhất, chúng được gọi là các vector hỗ trợ (support vector). Các đường nét đứt mà các support vector nằm trên đó được gọi là lề (margin).



Hình 29: Thuật toán SVM dùng để phân lớp 2 tập  $x_+$  và  $x_-$

SVM thực chất là một bài toán tối ưu, mục tiêu của thuật toán này là tìm được một không gian F và siêu phẳng sao cho sai số phân loại là thấp nhất.

Cho tập mẫu  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$  với  $x_i \in \mathbb{R}^n$ , và nhãn  $y_j \in \{-1, 1\}$  là nhãn lớp tương ứng của các vector  $x_i$ .

Ta có phương trình siêu phẳng chứa vector  $\vec{x}_i$  trong không gian:

$$\vec{x}_i \vec{w} + b = 0 \quad \text{Đặt } f(\vec{x}_i) = \text{sign}(\vec{x}_i \vec{w} + b) = \begin{cases} +1 & \text{với } \vec{x}_i \vec{w} + b > 0 \\ -1 & \text{với } \vec{x}_i \vec{w} + b < 0 \end{cases}$$

Như vậy,  $f(\vec{x}_i)$  biểu diễn sự phân lớp của vector  $\vec{x}_i$  vào 2 lớp cần phân loại.

Tức là  $\vec{y}_i = +1$  nếu  $\vec{x}_i$  thuộc lớp thứ nhất và  $\vec{y}_i = -1$  nếu  $\vec{x}_i$  thuộc lớp thứ hai.

#### ▪ Bài toán phân 2 lớp với SVM

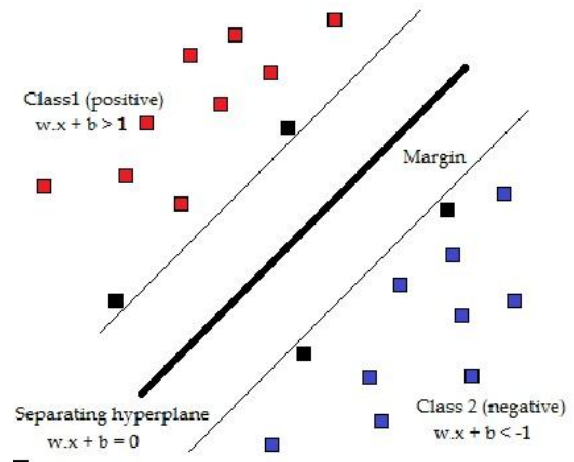
Bài toán đặt ra là: Xác định hàm phân lớp để phân các mẫu trong tương lai, nghĩa là với một mẫu dữ liệu mới  $\vec{x}_i$  thì cần được phải xác định  $\vec{x}_i$  được phân loại vào lớp +1 hay lớp -1.

Ta xét 3 trường hợp, mỗi trường hợp sẽ có một bài toán tối ưu, giải được bài toán tối ưu đó ta sẽ có được siêu phẳng cần tìm.

Trường hợp 1: Tập D có thể phân chia tuyến tính được mà không có nhiễu (tất cả các điểm được gán nhãn +1 thuộc về phía dương của siêu phẳng, tất cả các điểm được gán nhãn -1 thuộc về phía âm của siêu phẳng).

Ta sẽ tìm siêu phẳng tách với  $w \in \mathbb{R}^n$  là vector trọng số,  $b \in \mathbb{R}$  là hệ số tự do sao cho:

$$\text{Đặt } f(x_i) = \text{sign}(x_i w^T + b) = \begin{cases} +1 & \text{với } y_j = +1 \\ -1 & \text{với } y_j = -1 \end{cases} \quad \forall (x_i, y_j) \in D$$



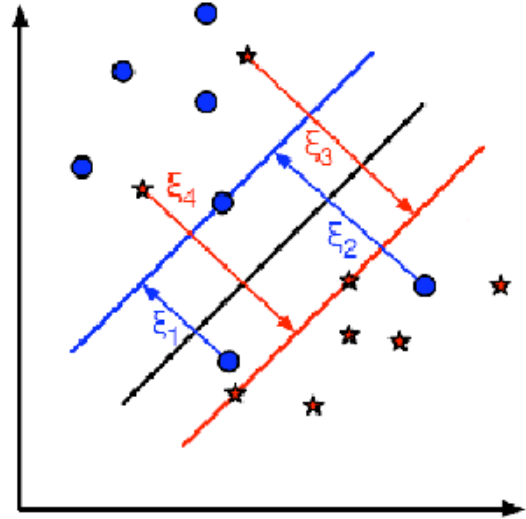
Hình 30: Tập dữ liệu được phân chia tuyến tính

Bài toán tối ưu: 
$$\begin{cases} \text{Min}(L(w)) = \frac{1}{2} \|w\|^2 \\ y_i(x_i \cdot w^T + b) \geq 1, i = 1 \dots l \end{cases}$$

Trường hợp 2: Tập dữ liệu D có thể phân chia tuyến tính được nhưng có nhiễu. Trong trường hợp này hầu hết các điểm được phân chia đúng bởi siêu phẳng. Tuy nhiên có một số điểm bị nhiễu như hình bên.

Trong trường hợp này ta sử dụng một lề mềm  $\varepsilon_i \geq 0$  sao cho:

$$y_i \cdot w^T + b \geq 1 - \varepsilon_i \text{ với } i = 1 \dots l$$



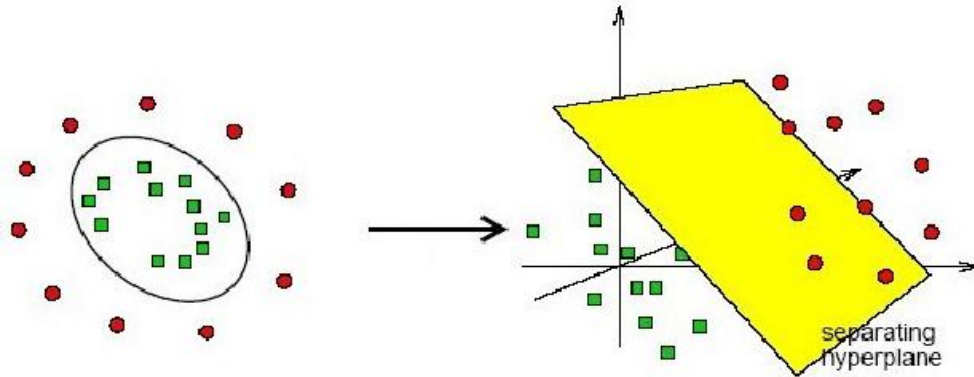
Hình 31: Tập dữ liệu phân chia tuyến tính nhưng có nhiễu.

Bài toán tối ưu:

$$\begin{cases} \text{Min}(L(w)) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \varepsilon_i \\ y_i(x_i \cdot w^T + b) \geq 1 - \varepsilon_i \text{ với } i = 1 \dots l \text{ và } \varepsilon_i \geq 0 \end{cases}$$

Trong đó C là tham số xác định trước, định nghĩa giá trị ràng buộc, C càng lớn thì mức độ lỗi thực nghiệm càng cao.

Trường hợp 3: Tập dữ liệu D không thể phân chia tuyến tính được, ta sẽ ánh xạ các vector dữ liệu x từ không gian n chiều vào một không gian m chiều (  $m > n$  ) sao cho trong không gian m chiều ta có thể phân chia tuyến tính được.



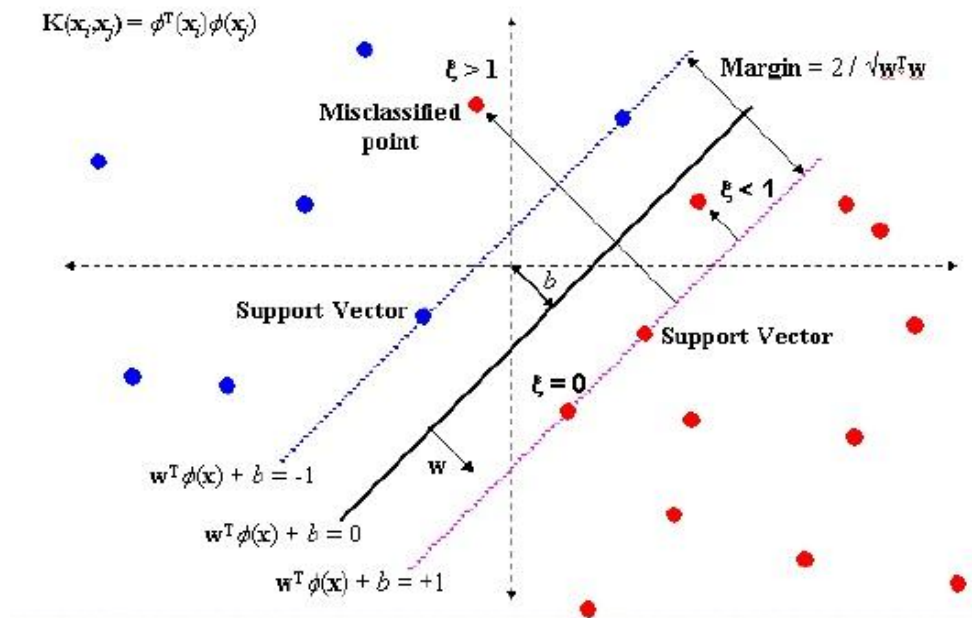
Hình 32: Tập dữ liệu không phân chia tuyến tính được.

Gọi  $\phi$  là một ánh xạ phi tuyến từ không gian  $\mathbb{R}^n$  vào không gian  $\mathbb{R}^m$ .

$$\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

Bài toán tối ưu: 
$$\begin{cases} \text{Min}(L(w, \varepsilon)) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \varepsilon_i \\ y_i(\phi(x_i) \cdot w^T + b) \geq 1 - \varepsilon_i \text{ với } i = 1 \dots l \text{ và } \varepsilon_i \geq 0 \end{cases}$$

Xét mô hình sau đây: trong không gian 2 chiều  $n = 2$ , tập dữ liệu được cho bởi các tập điểm trong mặt phẳng.



Hình 33: Biểu diễn tập dữ liệu trên không gian 2 chiều.

Ta tiến hành tìm siêu phẳng phân lớp dựa trên phương pháp SVM. Ta tìm 2 siêu phẳng phân tách (nét đứt màu xanh và màu hồng) sao cho khoảng cách giữa chúng là lớn nhất để có thể phân tách lớp này thành 2 phía. Các điểm nằm trên 2 siêu phẳng này là các vector hỗ trợ, chúng quyết định hàm phân tách dữ liệu. Siêu phẳng cần tìm là siêu phẳng nét đậm màu đen nằm giữa 2 siêu phẳng cần tách.

▪ **Bài toán phân đa lớp với SVM**

Để phân đa lớp thì kỹ thuật SVM sẽ chia không gian dữ liệu thành 2 thành phần và tiếp tục với không gian đã được phân chia. Khi đó hàm quyết định phân dữ liệu vào lớp thứ  $i$  sẽ là:  $f_i(x) = w_i^T x + b_i$

Những phần tử  $x$  là support vector nếu thỏa điều kiện:

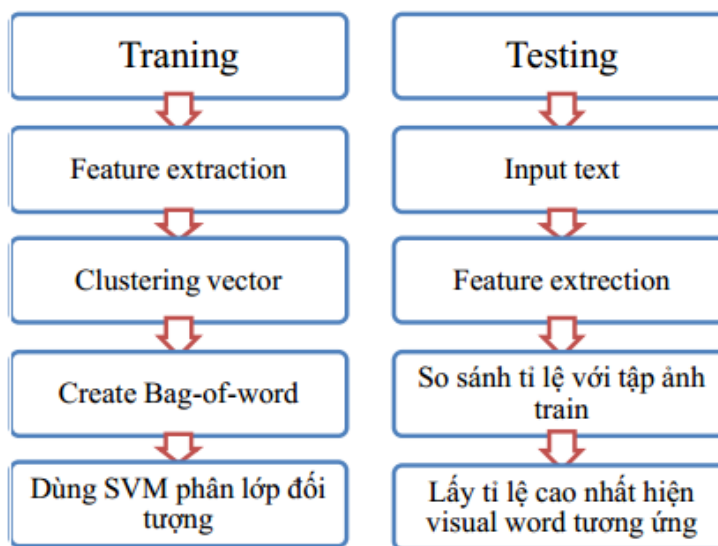
$$f_i(x) = \begin{cases} 1, & \in i \\ -1, & \notin i \end{cases}$$

Giả sử bài toán phân loại  $k$  lớp ( $k \geq 2$ ), ta sẽ tiến hành  $\frac{k(k-1)}{2}$  lần phân lớp nhị phân sử dụng phương pháp SVM. Mỗi lớp tiến hành phân tách với  $k - 1$  lớp còn lại để xác định  $k-1$  hàm phân tách (chiến lược “một đối một” (one against one)).

Kỹ thuật phân đa lớp bằng SVM hiện đang được tiếp tục nghiên cứu và phát triển.

## CHƯƠNG 5: MÔ TẢ BÀI TOÁN PHÂN LOẠI QUẦN ÁO

### 5.1 Mô hình bài toán



Hình 34: Mô hình phân loại quần áo.

### 5.2 Mô tả bài toán phân loại quần áo

- **Các bước để train tập ảnh**

Bước 1: Nhận một tập ảnh đầu vào

Tập ảnh đầu vào là tập dữ liệu gồm 15 nhãn khác nhau được mô tả tại mục 3.1, mỗi nhãn gồm 40 tấm hình được đưa vào để huấn luyện.

Bước 2: Rút trích đặc trưng ảnh thành các vector đặc trưng bằng giải thuật SURF

Thuật toán SUFT được sử dụng để rút trích những vector đặc trưng quan trọng dựa trên không gian tỉ lệ, và sử dụng kết quả xấp xỉ trên ma trận Hessian và ảnh tích hợp.

Bước 3: Gom cụm các vector đặc trưng vừa rút trích được bằng giải thuật k-means.

Trong demo bài này số cụm  $k = 500$  sau khi tiến hành các bước được mô tả ở mục 4.1, ta sẽ tạo được một tập visual word vocabulary tại bước 4.

Bước 4: Tạo ra visual word vocabulary tương ứng ở các cụm vừa gom được vào trong Bag-of-words. Sau đó, đếm số visual word của ảnh và biểu diễn thành dạng biểu đồ thể hiện tần số xuất hiện của visual word.

Bước 5: Dùng giải thuật Support Vector Machine để phân lớp cho đối tượng của bài toán

- ***Các bước để test tập ảnh***

Bước 1: Nhận vào tập một ảnh

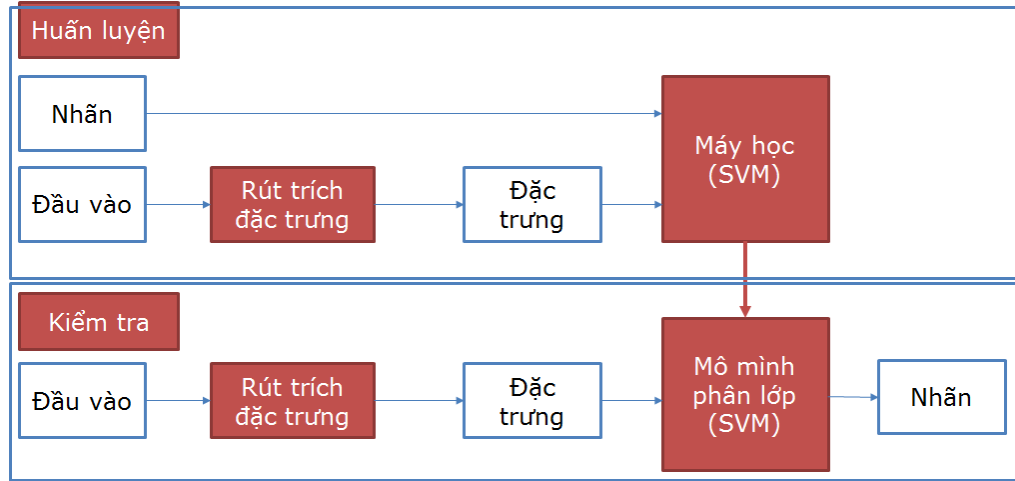
Bước 2: Rút trích đặc trưng của ảnh vừa nhận vào giống như rút trích đặc trưng khi train dữ liệu.

Bước 3: So sánh các đặc trưng ảnh vừa rút trích với tập train.

Bước 4: Tính khoảng cách giữa 2 vector đặc trưng gần nhất là visual word tương ứng là lớp cần được phân loại.

### **5.3 Phân lớp**

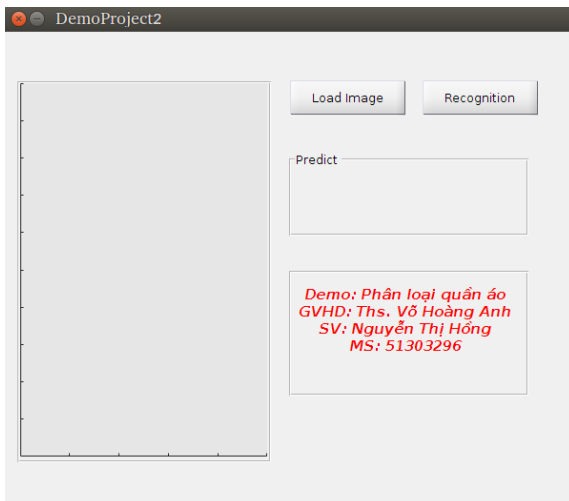
Nhiệm vụ của quá trình phân lớp một hệ thống nhận dạng là sử dụng các véc tơ đặc trưng được cung cấp bởi rút trích đặc trưng để xây dựng một mô hình nhận dạng đối tượng để khi đưa một đối tượng mới vào mô hình có thể dự đoán được lớp của đối tượng đó. Quá trình phân lớp sử dụng thuật toán SVM cài đặt.



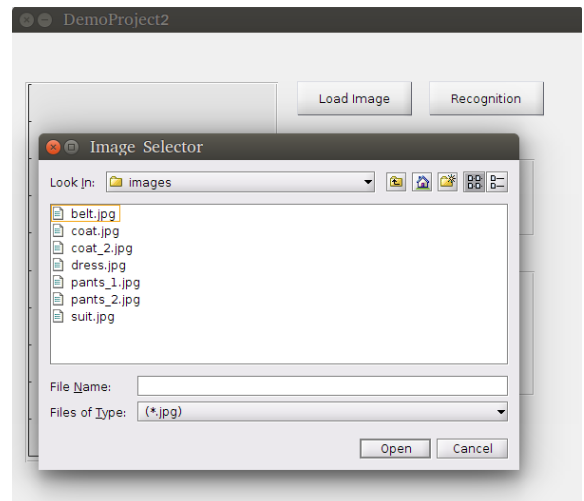
Hình 35: Mô hình quá trình phân lớp

Hình trên mô tả quá trình phân lớp. Bộ cơ sở dữ liệu được chia thành 2 tập huấn luyện và kiểm tra. Các chia được mô tả chi tiết trong mục 5.2. Tập huấn luyện sẽ là đầu vào của quá trình huấn luyện. Các đầu vào này sẽ được gán nhãn của lớp trước khi đưa vào cho máy học. Giải thuật huấn luyện SVM sẽ xây dựng một mô hình nhận dạng. Quá trình kiểm tra sẽ sử dụng mô hình nhận dạng này để dự đoán nhãn của tập kiểm tra đầu vào và gán nhãn cho dữ liệu.

## 5.4 Demo ứng dụng

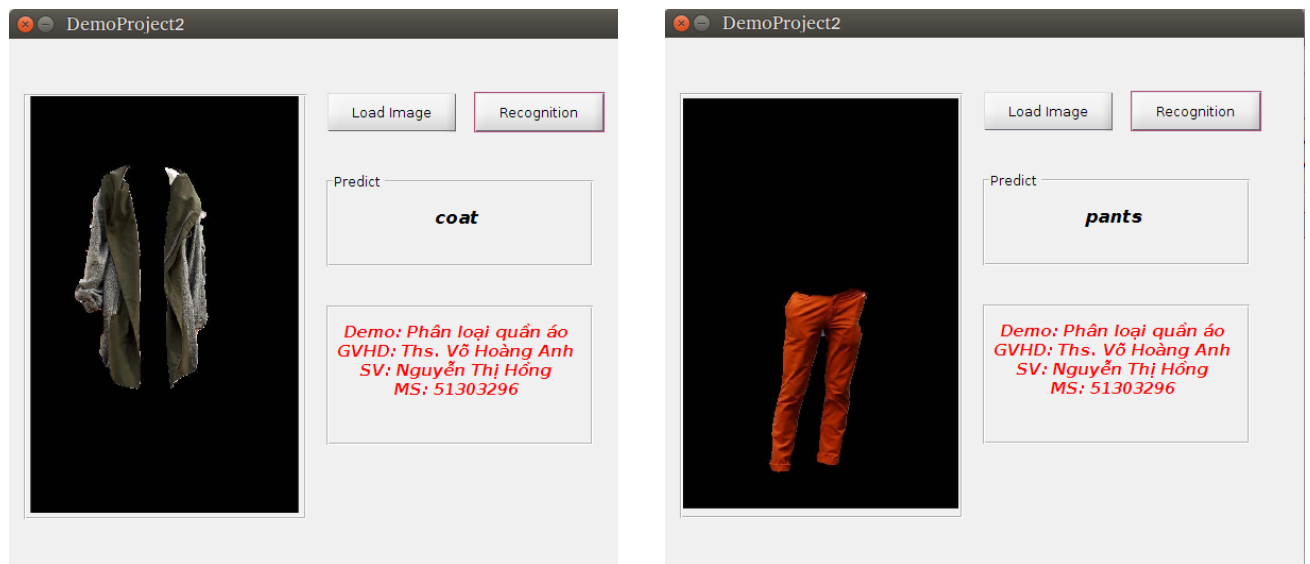


Hình 36: Giao diện demo ứng dụng.

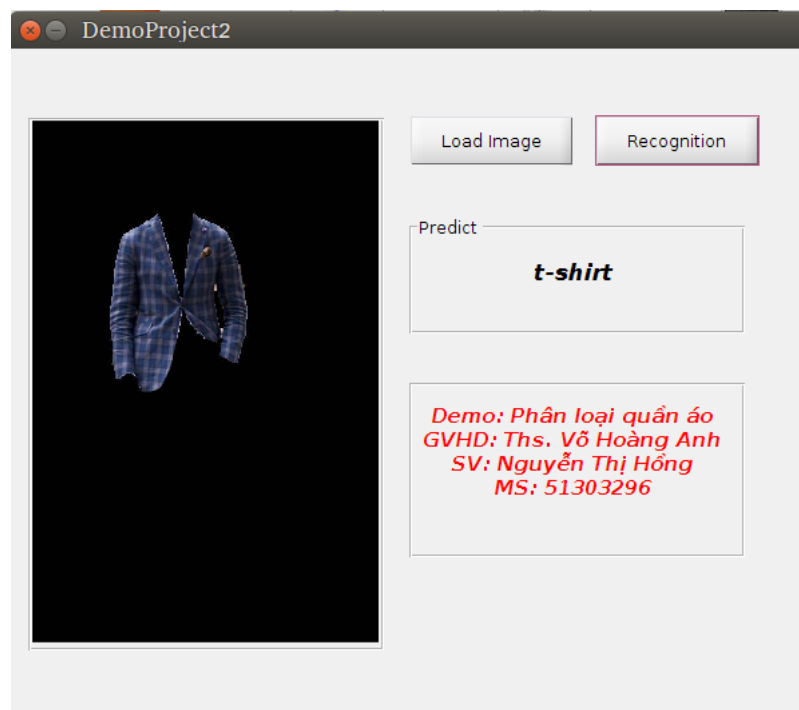


Hình 37: Chọn ảnh cần phân loại.





Hình 38: Nhãn cần phân loại được nhận dạng đúng.



Hình 39: Nhãn cần phân loại nhận dạng bị sai từ “Suit” thành “T-shirt”.

## CHƯƠNG 6: ĐÁNH GIÁ VÀ KẾT QUẢ THỰC NGHIỆM

### 6.1 Cách đánh giá

Quá trình huấn luyện và kiểm tra đánh giá độ chính xác của mô hình nhận dạng được xây dựng. Chia tập cơ sở dữ liệu thành 2 tập huấn luyện và kiểm tra để làm đầu vào cho 2 quá trình huấn luyện và kiểm tra theo phương pháp đánh giá chéo (cross validation) theo tỷ lệ 90:10 được sử dụng phổ biến hiện nay. Sử dụng phương pháp đánh giá chéo này nhằm tránh tình trạng quá khớp dữ liệu (over-fitting).

Tập cơ sở dữ liệu được chia đều thành mười phần bằng nhau tương ứng với chín phần của tập huấn luyện và một phần của tập kiểm tra cho một trong mười lần huấn luyện và kiểm tra. Tức là, mỗi lần huấn luyện và kiểm tra, chọn một phần dùng làm tập kiểm tra và chín phần còn lại dùng làm tập huấn luyện sao cho trong mỗi lần huấn luyện và kiểm tra tập kiểm tra không trùng với tập huấn luyện và trong mười lần huấn luyện và kiểm tra các tập kiểm tra là không trùng nhau.

Độ chính xác của mô hình nhận dạng được đánh giá theo công thức:

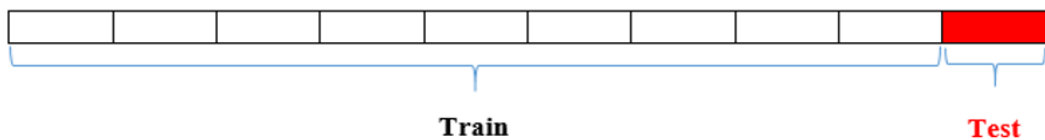
$$\text{Độ chính xác } (P) = \frac{\text{số hình nhận diện đúng} * 100}{\text{Tổng số hình nhận diện}}$$

Lần huấn luyện - kiểm tra 1



•  
•  
•

Lần huấn luyện - kiểm tra 10



Hình 40: Mô tả cách chia tập huấn luyện và tập kiểm tra.

## 6.2 Kết quả thực nghiệm

Sau khi hoàn thành các bước để nhận dạng đối tượng trong tập dữ liệu được mô tả ở trên, sử dụng thuật toán SURF để rút trích đặc trưng, dùng giải thuật k-mean để gom cụm và cuối cùng dùng SVM phân lớp. Mô hình nhận dạng được hiện thực trên máy có cấu hình core i7, 16GB Ram. Sau mười lần huấn luyện và kiểm tra ta thu được bảng thống kê kết quả như sau:

Lần	Số lượng hình training	Số lượng hình test	Tỉ lệ đúng (%)
1	45x15	5x15	24
2	45x15	5x15	25
3	45x15	5x15	37
4	45x15	5x15	28
5	45x15	5x15	33
6	45x15	5x15	35
7	45x15	5x15	28
8	45x15	5x15	29
9	45x15	5x15	29
10	45x15	5x15	27

Bảng 5: Kết quả thực nghiệm sau 10 lần chạy.

Tỉ lệ trung bình: **29.5 %**.

## **CHƯƠNG 7: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

### **7.1 Kết luận**

Các công việc được thực hiện trong đồ án:

- Tìm kiếm tập cơ sở dữ liệu cho bài toán.
- Xử lý hình ảnh dùng để làm tập huấn luyện và tập train.
- Cho máy học phân loại quần áo.
- Rút trích đặc trưng ảnh và phân loại ảnh.
- Demo ứng dụng.

#### **Khó khăn và hạn chế**

Các hạn chế trong đồ án như sau:

- Do tập dữ liệu hạn chế nên độ chính xác còn chưa cao so với các nghiên cứu khác.
- Một số khó khăn trong lĩnh vực thị giác máy tính.
- Để train và test tập dữ liệu thì yêu cầu máy tính phải có cấu hình cao, bộ nhớ lớn và có cài matlab.

### **7.2 Hướng phát triển**

Hiện nay, việc áp dụng các ứng dụng trong công nghệ thông tin vào đời sống là điều vô cùng quan trọng. Với nhu cầu con người càng ngày càng cao. Việc áp dụng bài toán phân loại quần áo trong việc phân loại và tìm kiếm trên các trang web như google là một hướng tìm hiểu và mở rộng bài toán này. Có thể phân loại quần áo trong các bài viết hay các tài liệu liên quan thông qua hình ảnh.

## TÀI LIỆU THAM KHẢO

### Tiếng Việt

Google search:

- Ảnh tích hợp
- Máy hỗ trợ vector
- Ma trận Hessian

### Tiếng Anh

[1]. Parsing Clothing in Fashion Photographs - Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, Tamara L Berg

<http://hi.cs.waseda.ac.jp/~esimo/publications/SimoSerraCVPR2015.pdf>.

[2]. A High Performance CRF Model for Clothes Parsing - Edgar Simo-Serra, Sanja Fidler, Francesc Moreno - Nogue, Raquel Urtasun

<http://hi.cs.waseda.ac.jp/~esimo/publications/SimoSerraCVPR2015.pdf>

[3]. Clothing Co-Parsing by Joint Image Segmentation and Labeling - Wei Yang, Ping Luo, Liang Lin

<https://pdfs.semanticscholar.org/3ce3/43da7581e017f9042ac9f45d7b542fe6f944.pdf>

[4]. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, “*Speeded Up Robust Features*” ETH Zurich, Katholieke Universiteit Leuven