### ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH ĐẠI HỌC CÔNG NGHỆ THÔNG TIN KHOA KHOA HOC MÁY TÍNH

SINH VIÊN THỰC HIỆN: NGUYỄN VĂN BIÊN - 10520245 PHAM DUY - 10520074

KHOÁ LUÂN TỐT NGHIỆP

# NGHIÊN CỨU KỸ THUẬT VÀ XÂY DỰNG ỨNG DỤNG TÌM KIẾM ĐỐI TƯỢNG TRÊN ẢNH



LUẬN VĂN CỬ NHÂN KHOA HỌC MÁY TÍNH

GIẢNG VIÊN HƯỚNG DẪN: TS. NGÔ ĐỨC THÀNH PGS. TS. LÊ ĐÌNH DUY

6, 2014

# LỜI CÁM ƠN

Tôi xin chân thành cám ơn ...

### TÓM TẮT

Trong những năm gần đây, truy vấn ảnh trên tập dữ liệu lớn là bài toán đang thu hút được nhiều sự quan tâm và có ý nghĩa quan trọng trong thực tiễn. Bài toán trên có thể phát biểu như sau: Đưa vào một hình ảnh có chứa đối tượng quan tâm và ngay lập tức trả về những hình ảnh có chứa đối tượng đó từ một tập dữ liệu trong thời gian thực. Các hệ thống truy vấn ảnh trên cơ sở dữ liệu lớn có nhiều ứng dụng quan trọng trong các lĩnh vực như nhận dạng đối tượng hay địa điểm, tìm kiếm video, phát hiện trùng lặp và tái tạo 3D, v.v... Tuy nhiên, bài toán trên cũng đang đối mặt với nhiều thách thức. Bên cạnh vấn đề về sự xuất hiện các biến thể của hình ảnh của đối tượng do sự khác nhau về độ sáng, kích thước, góc chụp hay bị che khuất một phần thì ở đây còn một vấn đề quan trọng khác là phải đảm bảo được thời gian thực hiện truy vấn đặc biệt là khi tìm kiếm trong tập dữ liệu lớn.

Rất nhiều công trình nghiên cứu đã được đề xuất để giải quyết vấn đề trên và đã đạt được nhiều bước tiến đáng chú ý. Hầu hết các công trình đó đều dựa trên mô hình Bag-of-Words (BoW), theo đó mỗi hình ảnh sẽ được biểu diễn bằng các đặc trưng cục bộ, sau đó các đặc trưng này được lượng tử hóa vào các visual word. Để tăng hiệu suất của quá trình truy vấn, người ta thường sử dụng mô hình Bag-of-Words kết hợp với phương pháp đánh chỉ mục ngược (Inverted Index). Thế nhưng cả Bag-of-Words và Inverted Index đều bỏ qua một thông tin quan trọng để tăng độ chính xác cho truy vấn, đó là thông tin không gian ảnh (spatial information) của các đặc trưng cục bộ.

Trong luận văn này, chúng tôi đề xuất một phương pháp nhằm tích hợp thông tin không gian ảnh vào phương pháp đánh chỉ mục ngược (Inverted Index) để nâng cao độ chính xác nhưng vẫn đảm bảo được

thời gian truy vấn nhanh. Kết quả thí nghiệm trên các tập dữ liệu chuẩn như Oxford 5k, Paris 6k và Holiday đã cho thấy tính hiệu quả của phương pháp này.

Từ khóa: Tìm kiếm ảnh - Image Search, Kích cỡ lớn - Large-Scale, Thông tin không gian - Spatial Information, Chỉ mục ngược - Inverted Index.

# Mục lục

M	lục lụ	ic i	v				
D	Danh sách hình vẽ Danh sách bảng						
D							
D	anh s	sách từ viết tắt vi	ii				
1	Giớ	i thiệu tổng quan	1				
	1.1	Mục tiêu và động lực chọn đề tài	1				
	1.2	Một vài hướng ứng dụng thực tế của hệ thống	2				
	1.3	Những thách thức đặt ra	3				
		1.3.1 Dây là subsection	4				
		1.3.1.1 Dây là sunsubsection	4				
		1.3.2 Đây là subsection tiếp theo	4				
	1.4	Kết chương	6				
2	Tổn	ng quan	7				
	2.1	First Section	7				
	2.2	Second Section	7				
		2.2.1 first subsection in the Second Section	8				
		2.2.2 second subsection in the Second Section	8				
		2.2.3 third subsection in the Second Section	8				
	2.3	Kết chương	8				

3	Các	tập d	ữ liệu và phương pháp đánh giá	10		
	3.1	First S	Section of the Third Chapter	10		
		3.1.1	first subsection in the First Section $\dots$	10		
		3.1.2	second subsection in the First Section	10		
			3.1.2.1 first subsub section in the second subsection	10		
		3.1.3	third subsection in the First Section	10		
			3.1.3.1 first subsub section in the third subsection	12		
			3.1.3.2 second subsub section in the third subsection	12		
	3.2	Second	d Section of the Third Chapter	12		
	3.3	Kết ch	nương	13		
4	Tícl	h hợp	thông tin không gian ảnh vào inverted index	14		
	4.1	First S	Section of the Third Chapter	14		
		4.1.1	first subsection in the First Section	14		
		4.1.2	second subsection in the First Section	14		
			4.1.2.1 $$ first subsub section in the second subsection	14		
		4.1.3	third subsection in the First Section	14		
			4.1.3.1 first subsub section in the third subsection	16		
			4.1.3.2 second subsub section in the third subsection	16		
	4.2	Second	d Section of the Third Chapter	16		
	4.3	Kết ch	nương	17		
5	Kết	luận '	và hướng phát triển	18		
Ρl	nụ lụ	c A		19		
Phụ lục B References						

# Danh sách hình vẽ

1.1	Airfoil Picture	Ę
3.1	TUI security program main features	11
4.1	TUI security program main features	15

# Danh sách bảng

# Danh mục từ viết tắt

 $\mathbf{BoW} \;\; \mathrm{Bag\text{-}of\text{-}Words}$ 

**SPM** Spatial Pyramid Matching

# Chương 1

# Giới thiệu tổng quan

### 1.1 Mục tiêu và động lực chọn đề tài

Trong những năm gần đây, cùng với sự phát triển của công nghệ thông tin, các lĩnh vực liên quan đến kỹ thuật số cũng đang có tốc độ phát triển chóng mặt. Các thiết bị kỹ thuật số như máy ảnh, máy quay phim kỹ thuật số, camera số, điện thoại di động có chức năng chụp hình, ... đang ngày càng phổ biến và không ngừng gia tăng về số lượng. Chính điều này đã làm sản sinh ra một lượng thông tin số khổng lồ bao gồm hình ảnh, video, v.v... Do đó, nhu cầu truy vấn thông tin từ kho dữ liệu hình ảnh, video ngày càng bức thiết hơn bao giờ hết.

Mục tiêu của luận văn này nhằm xây dựng một hệ thống truy vấn ảnh trên tập dữ liệu lớn, trong đó quá trình truy vấn hoàn toàn dựa trên nội dung của ảnh và kết quả phải được trả về gần như ngay lập tức với cơ sở dữ liệu gồm hàng triệu hình ảnh chưa được gán nhãn. Hệ thống này tập trung vào giải quyết vấn đề về tìm kiếm một đối tượng cụ thể như một địa điểm, một bức tranh, một bìa sách, v.v... Những đối tượng này có thể được chụp trong các điều kiện khác nhau như góc chụp, ánh sáng, kích thước hay bị che khuất. Do đó mục đích của hệ thống không phải là trả về những hình ảnh chụp gần giống nhau như chụp trong cùng một khung cảnh mà là trả về những hình ảnh có chứa đối tượng cần tìm. Ví dụ như khi đưa vào một bức hình có chứa Nhà thờ Đức Bà, kết quả trả về sẽ những bức hình có chứa nhà thờ Đức Bà chứ không phải trả về những nhà thờ có kiến trúc gần giống với Nhà thờ Đức Bà.

Những hệ thống truy vấn ảnh trên tập dữ liệu lớn có rất nhiều ứng dụng trong thực tế. Chúng tôi sẽ liệt kê sơ lược một vài ứng dụng trong phần tiếp theo.

# 1.2 Một vài hướng ứng dụng thực tế của hệ thống

Trong cuộc sống, ta có thể dễ dàng bắt gặp những ứng dụng vô cùng hữu ích của các hệ thống truy vấn đối tượng trên ảnh. Dưới đây là một vài hướng ứng dụng cu thể:

Nhận dạng đối tượng, sản phẩm. Với sự phổ biến của điện thoại thông minh và internet, một người có thể dễ dàng dùng điện thoại chụp một tấm hình và hỏi hệ thống về thông tin của đối tượng trong tấm hình đó. Ví dụ, tại một cửa hàng, một người mua hàng có thể tham khảo giá của một sản phẩm tại các cửa hàng khác; trong thư viện, một độc giả có thể tìm được những cuốn sách nào chứa hình ảnh mình quan tâm; khi đi thăm bảo tàng, du khách có thể tìm kiếm thêm thông tin về một hiện vật trong đó, v.v...

Nhận dạng địa điểm. Vị trí địa lý của nơi chụp tấm hình cũng có thể được xác định bằng việc truy vấn thông tin của đối tượng trong hình từ những cơ sở dữ liệu lớn chứa hình ảnh và thông tin vị trí như Google Street View hay kho hình ảnh có lưu kèm thông tin GPS. Hệ thống này có thể là một giải pháp thay thế rẻ tiền cho các thiết bị có GPS. Chẳng hạn, khi một du khách đến một nơi mà anh ta chưa bao giờ đặt chân tới nhưng lại không GPS hay bản đồ, anh ta có thể chụp một tấm hình của một tòa nhà hay những cảnh tại nơi đó để xác định được vị trí chính xác của mình.

Tìm kiếm và quản lý kho dữ liệu video. Hàng ngày, một lượng lớn dữ liệu video được sinh ra và ta không thể nào quản lý hết được nội dung của chúng. Ví dụ, một hãng phim hay truyền hình muốn tìm kiếm tắt cả các đoạn quảng cáo có liên quan đến một nhãn hiệu sản phẩm mà họ đã từng phát trong vài năm gần đây, một hệ thống truy vấn ảnh sẽ dễ dàng thực hiện điều này chỉ với một bức hình của sản phẩm.

Gán nhãn ảnh tư đông. Những tấm ảnh có thể được gán nhãn một cách tự

động về địa điểm hay đối tượng trong hình để dễ dàng cho việc tìm kiếm và quản lý sau này. Ví dụ, người dùng có thể dễ dàng tìm kiếm được những bức hình chụp tại một địa điểm nào đó mà không cần biết nó nằm trong album nào hay được chụp ngày nào. Những hệ thống lớn lưu trữ ảnh lớn như của Facebook có thể dễ dàng phát hiện và gán nhãn khuôn mặt người nhưng vẫn chưa thể nhận dạng được địa điểm mà tấm hình được chụp từ nội dung chứa trong hình.

Sử dụng trong quảng cáo theo ngữ cảnh. Rất nhiều công ty quảng cáo đặt màn hình tại nơi công cộng để quảng cáo cho các sản phẩm của mình nhưng các quảng cáo này chưa thực sự hướng người dùng và kém hiệu quả. Việc sử dụng một hệ thống có thể quảng cáo theo ngữ cảnh và hướng đúng đối tượng người dùng sẽ giúp việc quảng cáo hiệu quả hơn. Ví dụ, một camera trong thang máy có thể tự động phát hiện được những sản phẩm người đi thang máy đang dùng như nhãn hiệu chai nước họ đang uống, nhãn hiệu quần áo họ đang mặc,... để lựa chọn được những quảng cáo phù hợp với đối tượng người dùng và phát trên màn hình.

Tăng tính tương tác thực tế. Với sự ra đời của các sản phẩm công nghệ gần gũi với cuộc sống như Google Glass, việc nhận dạng đối tượng trong thời gian thực sẽ mang đến nhiều thông tin hữu ích cho người dùng.

Hỗ trợ cho các hệ thống thị giác máy tính khác. Hệ thống truy vấn đối tượng có thể được dùng để hỗ trợ cho các hệ thống thị giác máy tính khác. Một ví dụ điển hình là hệ thống tự động tái tạo hình ảnh ba chiều sẽ cần gom cụm các hình ảnh của cùng một đối tượng từ một tập dữ liệu lớn.

#### 1.3 Những thách thức đặt ra

Để giải quyết bài toán truy vấn đối tượng trên tập dữ liệu ảnh lớn, có rất nhiều thách thức được đặt ra. Dưới đây chúng tôi sẽ trình bày một vài thách thức trong bài toán này:

Sự biến đổi bề ngoài của đối tượng trong hình ảnh. Một hệ thống truy vấn đối tượng trên hình ảnh cần phải trả về được các hình ảnh có chứa đối tượng quan tâm bất chấp mọi thay đổi trên bên ngoài của đối tượng. Những thay đổi đó có thể đến từ rất nhiều nguyên nhân khác nhau. Đó có thể do tác động từ các

yếu tố bên ngoài khi chụp hình như điều kiện chiếu sáng, góc chụp của camera hay những tùy chỉnh khác nhau của các camera về độ tương phản, độ phân giải, màu sắc,... Cùng với đó là những hình ảnh của đối tượng được chụp với góc xoay, kích thước hình hay tỉ lệ khác nhau. Hoặc xa hơn nữa là những trường hợp đối tượng bị che khuất, cắt ghép, v.v... Một vài dạng thay đổi kể trên được thể hiện qua hình XXXX. Có một trường hợp khác là do những thay đổi từ chính bản thân đối tượng do các điều kiện bên ngoài ví dụ như đối tượng bị cũ đi hay bị xuống cấp theo thời gian.

Các loại đặc tính vật lý khác nhau trên mỗi đối tượng. Dựa trên các đặc tính vật lý người ta chia đối tượng thành các loại khác nhau. Có những đối tượng mà đặc tính thể hiện rõ nét nhất qua cấu trúc bề mặt, nhưng có cái lại qua màu sắc hay hình dạng, v.v... Ví dụ như với những con bướm, đặc trưng cho chúng không phải là hình dạng, kích cỡ vì đa phần các loài bướm đều có hình dạng, kích cỡ gần giống nhau mà ở đây là các họa tiết, màu sắc trên cánh bướm; Hay với những loại lá cây thì đặc trưng về màu sắc, họa tiết lại không cung cấp nhiều thông tin bằng hình dạng của lá.

Kích cỡ của tập dữ liệu lớn. Tập dữ liệu hình ảnh lớn thường bao gồm hàng triệu bức ảnh, vậy nên để người dùng có thể tương tác trực tiếp với hệ thống thông qua một thiết bị phía client như điện thoại di động thì đòi hỏi truy vấn phải được trả về trong thời gian ngắn chấp nhận được. Do đó cần phải có một thuật toán nhận dạng hiệu quả, chi phí thấp. Đồng thời những hình ảnh cũng phải được xử lý để lưu trữ sao cho tiết kiệm nhất để phù hợp với kích cỡ của RAM vì nếu lưu trữ trên ổ cứng sẽ mất rất nhiều thời gian để truy xuất và không thể đạt được yêu cầu về thời gian.

#### 1.3.1 Dây là subsection

#### 1.3.1.1 Dây là sunsubsection

Hạn chế dùng đến x.x.x....

#### 1.3.2 Đây là subsection tiếp theo

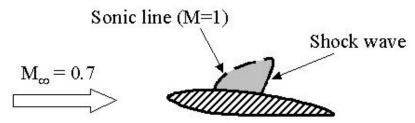
... and some more ...

Now I would like to cite the following: and ?.

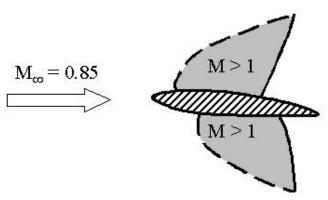
I would also like to include a picture ...



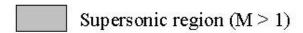
a) Subsonic flow



b) Low transonic Mach number



c) High transonic Mach number



Hình 1.1: Airfoil Picture

So as we have now labelled it we can reference it, like so (1.1) and it is on

Page 5. And as we can see, it is a very nice picture and we can talk about it all we want and when we are tired we can move on to the next chapter ...

I would also like to add an extra bookmark in acroread like so ...

### 1.4 Kết chương

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

# Chương 2

# Tổng quan

#### 2.1 First Section

iết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 2.2 Second Section

iết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 2.2.1 first subsection in the Second Section

... and some more ... Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 2.2.2 second subsection in the Second Section

... and some more ... Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 2.2.3 third subsection in the Second Section

 $\dots$  and some more  $\dots$ 

### 2.3 Kết chương

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX. Viết luận văn bằng LATEX. Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

# Chương 3

# Các tập dữ liệu và phương pháp đánh giá

### 3.1 First Section of the Third Chapter

And now I begin my third chapter here ...

#### 3.1.1 first subsection in the First Section

... and some more

#### 3.1.2 second subsection in the First Section

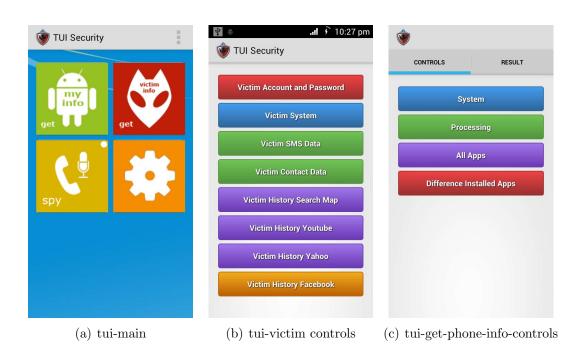
... and some more ...

#### 3.1.2.1 first subsub section in the second subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

#### 3.1.3 third subsection in the First Section

... and some more ...



Hình 3.1: TUI security program main features

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX. Viết luận văn bằng

LĂTEX. Viết luận văn bằng LĂTEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 3.1.3.1 first subsub section in the third subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more ...

#### 3.1.3.2 second subsub section in the third subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

### 3.2 Second Section of the Third Chapter

and here I write more ...

### 3.3 Kết chương

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

# Chương 4

# Tích hợp thông tin không gian ảnh vào inverted index

### 4.1 First Section of the Third Chapter

And now I begin my third chapter here ...

#### 4.1.1 first subsection in the First Section

... and some more

#### 4.1.2 second subsection in the First Section

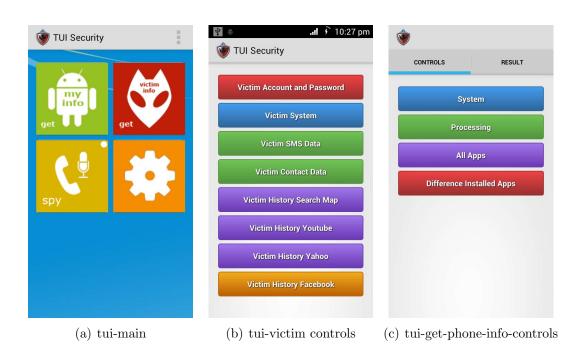
... and some more ...

#### 4.1.2.1 first subsub section in the second subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

#### 4.1.3 third subsection in the First Section

... and some more ...



Hình 4.1: TUI security program main features

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX. Viết luận văn bằng

ETEX. Viết luận văn bằng ETEX.

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

#### 4.1.3.1 first subsub section in the third subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more ...

#### 4.1.3.2 second subsub section in the third subsection

... and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it ...

#### 4.2 Second Section of the Third Chapter

and here I write more ...

### 4.3 Kết chương

Viết luận văn bằng LATEX. Viết luận văn bằng LATEX.

### Chương 5

# Kết luận và hướng phát triển

Here I put my conclusions ... Viết luận văn bằng LATEX. Viết luận văn

# Phụ lục A

Đây là phụ lục A

# Phụ lục B

Đây là phụ lục B

### References

- AHAMED, S.I., HAQUE, M.M., HOQUE, M.E., RAHMAN, F. & TALUKDER, N. (2010). Design, analysis, and deployment of omnipresent formal trust model (ftm) with trust bootstrapping for pervasive environments. *Journal of Systems and Software*, 83, 253 270.
- Andress, M. (2002). Wireless lan security. *Information Systems Security*, **11**, 29–33.
- BARBEAU, M. (2012). Point-to-point voice over ad hoc networks: A survey. *Pervasive and Mobile Computing*, **8**, 376 387.
- Bennett, D. (2012). The challenges facing computer forensics investigators in obtaining information from mobile devices for use in criminal investigations. *Information Security Journal: A Global Perspective*, **21**, 159–168.
- Bertino, E. & Kirkpatrick, M. (2009). Location-aware authentication and access control concepts and issues. In *Advanced Information Networking and Applications*, 2009. AINA '09. International Conference on, 10–15.
- Breu, R., Burger, K., Hafner, M. & Popp, G. (2004). Towards a systematic development of secure systems. *Information Systems Security*, **13**, 5–13.
- Bunting, S. (2012). EnCase Computer Forensics The Official EnCE: EnCase Certified Examiner Study Guide. John Wiley & Sons.