

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ
TRUYỀN THÔNG**

Nguyễn Thị Bích Hạnh

**NGHIÊN CỨU KỸ THUẬT TRA CỨU ẢNH
VÀ ỨNG DỤNG TRONG TRA CỨU CÂY THUỐC**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên – 2013

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ
TRUYỀN THÔNG**

Nguyễn Thị Bích Hạnh

**NGHIÊN CỨU KỸ THUẬT TRA CỨU ẢNH
VÀ ỨNG DỤNG TRONG TRA CỨU CÂY THUỐC**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. Nguyễn Văn Tảo

Thái Nguyên - 2013

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn với tên đề tài “*Nghiên cứu kỹ thuật tra cứu cứu ảnh và ứng dụng trong tra cứu cây thuốc*” là kết quả nghiên cứu của bản thân học viên, dựa trên nghiên cứu và thu thập cơ sở dữ liệu thực tiễn dưới sự hướng dẫn khoa học của TS. Nguyễn Văn Tảo. Các số liệu, kết quả nêu trong luận văn là hoàn toàn trung thực, chưa từng được công bố dưới bất cứ hình thức nào.

Thái Nguyên, ngày 15 tháng 7 năm 2013

Người cam đoan

Nguyễn Thị Bích Hạnh

LỜI CẢM ƠN

Trong suốt thời gian hoàn thành luận văn nghiên cứu, tôi luôn nhận được giúp đỡ tận tình của thầy giáo hướng dẫn và biết bao người thân yêu bên tôi.

Nhân dịp này, tôi xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới TS. Nguyễn Văn Tảo về những chỉ dẫn khoa học, những góp ý quý báu và tận tình hướng dẫn tôi trong suốt quá trình thực hiện luận văn tốt nghiệp.

Tôi xin chân thành cảm ơn tập thể các Thầy, Cô giáo trong trường Đại học Công nghệ thông tin và Truyền thông đã luôn nhiệt tình quan tâm, giúp đỡ và tạo mọi điều kiện tốt nhất cho tôi trong suốt quá trình học tập tại trường.

Xin chân thành cảm ơn tới tập thể các thầy cô và các bạn đồng nghiệp Trung tâm Nghiên cứu và Phát triển Nông Lâm nghiệp miền núi phía Bắc - Trường Đại học Nông Lâm nơi tôi đang công tác; các bạn trường Đại học Y Thái Nguyên đã luôn cổ vũ, động viên và tạo mọi điều kiện thuận lợi để tôi hoàn thành luận văn của mình.

Cuối cùng, tôi muốn bày tỏ lòng cảm ơn vô hạn tới bạn bè, tới những người thân yêu luôn bên tôi. Xin bày tỏ lòng cảm ơn sâu sắc tới bố mẹ, chồng, con và những người thân trong gia đình đã là điểm tựa tinh thần lẫn vật chất cho tôi trong những lúc khó khăn, luôn bên cạnh và động viên tôi trong suốt quá trình thực hiện luận văn tốt nghiệp.

Xin trân trọng cảm ơn tất cả sự giúp đỡ quý báu đó!

Sinh viên

Nguyễn Thị Bích Hạnh

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC HÌNH.....	vi
DANH MỤC CÁC BẢNG.....	vii
PHẦN MỞ ĐẦU.....	1
Chương 1. TỔNG QUAN VỀ TRA CỨU ẢNH DỰA VÀO NỘI DUNG.....	3
1.1. Tổng quan về xử lý ảnh.....	3
1.2. Khái quát về tra cứu ảnh	5
1.2.1. Đặc trưng màu	7
1.2.2. Đặc trưng màu vân	9
1.2.3. Đặc trưng về hình dạng	9
1.2.4. Thông tin không gian.....	9
1.2.5. Phân đoạn.....	11
1.3. Các chức năng cơ bản của một hệ thống tra cứu ảnh dựa vào nội dung	12
1.4. Những ứng dụng cơ bản của tra cứu ảnh dựa vào nội dung.....	15
1.5. Một số hệ thống tra cứu ảnh theo nội dung tiêu biểu	16
1.5.1. Hệ thống QBIC	16
1.5.2. Hệ thống Photobook	16
1.5.3. Hệ thống VisualSEEK và WebSEEK	16
1.5.4. Hệ thống RetrievalWare	17
1.5.5. Hệ thống Imatch.....	17
1.5.6. Mô tơ tìm kiếm WWW ImageRover	17
Chương 2. MỘT SỐ KỸ THUẬT TRA CỨU ẢNH DỰA VÀO NỘI DUNG	18
2.1. Tra cứu ảnh dựa vào màu sắc	18
2.1.1. Màu sắc.....	18
2.1.2. Một số đặc tính đặc biệt về màu sắc.....	19
2.1.3. Các không gian màu	19
2.1.4. Lượng tử hóa màu và lược đồ màu.....	24

2.1.5. Đo khoảng cách giữa các lược đồ màu	29
2.2. Tra cứu ảnh dựa vào hình dạng	32
2.2.1. Khái niệm về hình dạng	32
2.2.2. Đặc điểm hình dạng với việc tìm kiếm ảnh.....	32
2.2.3. Lược đồ hình dạng	34
2.2.4. Biểu diễn hình dạng trên cơ sở vùng và thước đo đồng dạng	35
2.3. Tra cứu ảnh dựa vào vân	39
2.3.1. Vân là gì?.....	39
2.3.2. Tra cứu ảnh dựa vào vân.....	40
2.3.3. Phân hoạch vùng nhị phân cục bộ.....	41
2.3.4. Phân đoạn vân	41
2.4. Phương pháp tra cứu ảnh sử dụng dấu hiệu nhị phân	41
2.4.1. Giới thiệu	41
2.4.2. Kiến trúc hệ thống	43
2.4.3. Kỹ thuật tra cứu ảnh theo nội dung dựa vào dấu hiệu nhị phân.....	44
2.4.4. Độ đo tương tự.....	48
Chương 3. XÂY DỰNG CHƯƠNG TRÌNH TRA CỨU ẢNH CÂY THUỐC.....	52
3.1. Giới thiệu bài toán tra cứu ảnh cây thuốc	52
3.2. Phân tích bài toán	53
3.3. Các biểu đồ	57
3.3.1 Biểu đồ Use Case.....	57
3.3.2. Biểu đồ trình tự và biểu đồ hoạt động.....	57
3.4. Thiết kế cơ sở dữ liệu	62
3.5. Xây dựng chương trình.....	63
3.5.1. Ngôn ngữ sử dụng.....	63
3.5.2. Giao diện chương trình.....	63
3.6. Thử nghiệm và đánh giá.....	68
PHẦN KẾT LUẬN.....	71
TÀI LIỆU THAM KHẢO.....	72

DANH MỤC CÁC CHỮ VIẾT TẮT

STT	Từ viết tắt	Từ viết đầy đủ	Diễn giải
1	VIR	Visual Information Retrieval	Tra cứu thông tin
2	CBIR	Content Based Image Retrieval	Tra cứu ảnh dựa vào nội dung
3	QBE	Query by Example	Truy vấn bởi ảnh mẫu
4	QBF	Query by Feature	Truy vấn bởi đặc trưng
5	QBIC	Query By Image Content	Truy vấn ảnh dựa vào nội dung
6	RGB	Red Green Blue	Đỏ, Xanh lục, Xanh lơ
7	HSV	Hue, Saturation, Value	Màu, độ bão hòa màu, độ sáng
8	CCV	Color Coherence Vectors	Véc tơ gắn kết màu
9	CIE	Commission International d'Éclairage	Ủy ban quốc tế về màu sắc
10	CMY	Cyan-Magenta-Yellow	Lục lam, đỏ tươi, vàng
11	CBA	Constant Bin Allocation	Phân bố bin liên tục
12	VBA	Variable Bin Allocation	Phân bố biến bin
13	GCH	Global Color Histogram	Lược đồ màu toàn cục
14	LCH	Local Color Histogram	Lược đồ màu cục bộ
15	GIS	Geographic Information System	Hệ thống thông tin địa lý
16	CSDL	Cơ sở dữ liệu	

DANH MỤC CÁC HÌNH

Hình 1.1: Các bước cơ bản trong một hệ thống xử lý ảnh	3
Hình 1.2: Kiến trúc hệ thống tra cứu ảnh dựa vào nội dung.....	13
Hình 2.1: Mô tả không gian màu RGB	20
Hình 2.2: Mô tả không gian màu HSV	21
Hình 2.3: Lược đồ màu HIS cải tiến.	28
Hình 2.4: Khoảng cách dạng Minkowski.....	29
Hình 2.5: Khoảng cách dạng toàn phương.....	30
Hình 2.6: Hình dạng sau khi đã chồng lưới	35
Hình 2.7: Hình dạng trên hình 2.6 sau khi đã được chuẩn hóa.....	37
Hình 2.8: Đo hai hình dạng trên cơ sở chỉ mục	37
Hình 2.9: Một số loại vân tiêu biểu.....	40
Hình 2.10: Kiến trúc của kỹ thuật tra cứu ảnh sử dụng dấu hiệu nhị phân.....	43
Hình 2.11: Mẫu thiết lập hình ảnh.	45
Hình 3.1: Kiến trúc tổng quát của hệ thống tra cứu ảnh.	54
Hình 3.2: Mô hình chi tiết của hệ thống tra cứu ảnh	55
Hình 3.3: Tiền xử lý ảnh dữ liệu.	57
Hình 3.4: Biểu đồ Use Case.	57
Hình 3.5: Biểu đồ trình tự của tác nhân tra cứu.	58
Hình 3.6: Biểu đồ cộng tác của tác nhân tra cứu.	58
Hình 3.7: Biểu đồ hoạt động của tác nhân tra cứu ảnh.	59
Hình 3.8: Biểu đồ trình tự của tác nhân quản lý CSDL ảnh.	60
Hình 3.9: Biểu đồ cộng tác của tác nhân quản lý CSDL ảnh.....	60
Hình 3.10: Biểu đồ hoạt động của chức năng quản lý cơ sở dữ liệu ảnh.....	61
Hình 3.11: Màn hình khởi tạo của chương trình.....	63
Hình 3.12: Màn hình chọn cơ sở dữ liệu ảnh.....	64
Hình 3.13: Màn hình chọn ảnh cần truy vấn 1	65
Hình 3.14: Màn hình kết quả đối với ảnh truy vấn 1	65
Hình 3.15: Danh sách ảnh hiển thị thông tin được sắp xếp với ảnh truy vấn 1	66
Hình 3.16: Màn hình chọn ảnh truy vấn 2	67
Hình 3.17: Màn hình kết quả đối với ảnh truy vấn 2	67
Hình 3.18: Danh sách ảnh hiển thị thông tin được sắp xếp với ảnh truy vấn 2	68

DANH MỤC CÁC BẢNG

Bảng 2.1: Dấu hiệu nhị phân của hình 2.9 sử dụng CBA	46
Bảng 2.2: Dấu hiệu nhị phân của hình 2.9 sử dụng VBA.....	47
Bảng 2.3: Chuỗi dấu hiệu nhị phân minh họa sự giống và khác nhau của ảnh	47
Bảng 3.1: Bảng cơ sở dữ liệu ảnh.	62
Bảng 3.2: Bảng cơ sở dữ liệu nhị phân.	62
Bảng 3.3: Bảng kết quả thực hiện truy vấn 1.....	62
Bảng 3.4: Bảng kết quả thực hiện truy vấn 2.....	70

PHẦN MỞ ĐẦU

Ngày nay cùng với sự phát triển của các ngành khoa học khác, ngành công nghệ thông tin cũng có những bước phát triển nhanh chóng, ngày càng có nhiều ứng dụng vào thực tiễn cuộc sống và đem lại hiệu quả cao. Việc quản lý thông tin rất cần tới các công cụ để thu thập, phân loại tổ chức ở các dạng dữ liệu khác nhau một cách hiệu quả. Trong khi các hệ thống quản lý thông tin dựa vào văn bản đã rất thành công nhưng chúng vẫn không đáp ứng được việc quản lý, khai thác và xử lý dữ liệu với khối lượng khổng lồ như hiện nay. Ảnh số ngày càng thu hút được sự quan tâm của nhiều người, một phần là do các thiết bị thu nhận ảnh số ngày càng phổ biến và có giá cả phù hợp, cho phép nhiều người có thể sở hữu và sử dụng. Tra cứu và quản trị cơ sở dữ liệu ảnh thu hút sự quan tâm của nhiều nhà nghiên cứu từ những năm 1970. Với sự tăng nhanh về tốc độ máy tính và giảm chi phí bộ nhớ, các cơ sở dữ liệu ảnh chứa hàng nghìn thậm chí hàng triệu ảnh được sử dụng trong nhiều lĩnh vực ứng dụng khác nhau như y học, ảnh vệ tinh, các cơ sở dữ liệu ảnh sinh học và các cơ sở dữ liệu ảnh phong cảnh... Các ứng dụng này đòi hỏi độ chính xác tra cứu cao.

Khi có nhu cầu tra cứu một vài bức ảnh nào đó trong một cơ sở dữ liệu có hàng trăm ngàn ảnh, điều này khó có thể thực hiện được khi ta tra cứu bằng tay theo cách thông thường, nghĩa là xem từng tấm ảnh một cho đến khi thấy ảnh có nội dung cần tìm. Với sự tăng nhanh về số lượng ảnh, cách tiếp cận tra cứu ảnh dựa vào chú thích ảnh thủ công trở nên không khả thi về cả thời gian và chi phí.

Do đó, khi số lượng ảnh được lưu trữ trở nên rất lớn thì vấn đề đặt ra là phải có những phương pháp tổ chức cơ sở dữ liệu ảnh tốt cùng với những kỹ thuật tra cứu, tra cứu ảnh hiệu quả, có độ chính xác cao và có hiệu năng tốt mới đáp ứng được nhu cầu ngày càng cao của con người. Việc xây dựng các hệ thống tra cứu ảnh là rất cần thiết. Trong thực tế, bài toán tra cứu ảnh số có rất nhiều ứng dụng quan trọng. Ví dụ trong lĩnh vực ngân hàng việc so sánh chữ ký của khách hàng với mẫu chữ ký đã được lưu trữ sẵn có thể thực hiện rất nhanh và chính xác nếu có được một phần