

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

NGUYỄN THANH TÂN

**ĐỀ XUẤT CẢI TIẾN THUẬT TOÁN TÌM KIẾM
TRONG CƠ SỞ DỮ LIỆU MỜ**

LUẬN VĂN THẠC SỸ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2016

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

NGUYỄN THANH TÂN

**ĐỀ XUẤT CẢI TIẾN THUẬT TOÁN TÌM KIẾM
TRONG CƠ SỞ DỮ LIỆU MỜ**

Chuyên ngành: Khoa học máy tính

Mã số: 60480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

PGS. TS. NGUYỄN THIỆN LUẬN

Thái Nguyên - 2016

LỜI CAM ĐOAN

Tôi xin cam đoan những nội dung trong luận văn “Đề xuất cải tiến thuật toán tìm kiếm trong cơ sở dữ liệu mờ” là do tôi thực hiện dưới sự hướng dẫn trực tiếp của thầy giáo **PGS. TS. Nguyễn Thiện Luận**.

Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng tên tác giả, tên công trình, thời gian, địa điểm công bố.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá, tôi xin chịu trách nhiệm hoàn toàn.

Thái Nguyên, ngày 12 tháng 5 năm 2016

Học viên

Nguyễn Thanh Tân

LỜI CẢM ƠN

Em xin bày tỏ lòng biết ơn sâu sắc tới **PGS. TS. Nguyễn Thiện Luận** người thầy kính mến đã hướng dẫn, chỉ dạy tận tình để em hoàn thành luận văn này. Em xin chân thành cảm ơn các thầy, cô giáo trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên, Khoa đào tạo sau đại học - Đại học Thái Nguyên, Viện Công nghệ thông tin Việt Nam đã đón nhận và truyền thụ kiến thức cho em trong suốt quá trình học tập vừa qua.

Tôi cũng xin cảm ơn cơ quan, bạn bè đồng nghiệp, gia đình và những người thân đã cùng chia sẻ, giúp đỡ, động viên và tạo mọi điều kiện thuận lợi để tôi hoàn thành nhiệm vụ học tập và bản luận văn này.

Thái Nguyên, ngày 12 tháng 5 năm 2016

Học viên

Nguyễn Thanh Tân

DANH MỤC TỪ TIẾNG ANH VÀ VIỆT TẮT

ANN	Approximative Nearest Neighbour
BK - trees	Cây Burkhard - Keller
Codd E. F.	Edgar Frank "Ted" Codd
Damerau - Levenshtein distance	Khoảng cách Damerau - Levenshtein
DoG	Difference of Gaussian
Episode	Một hồi
Full text search	Tìm kiếm toàn văn
Fuzzy Databases	Cơ sở dữ liệu mờ
Fuzzy join	Phép kết nối mờ
Fuzzy Natural joint	Phép kết nối tự nhiên mờ
Fuzzy projection	Phép chiếu mờ
Fuzzy Search	Tìm kiếm mờ
Fuzzy Selection	Phép chọn mờ
Fuzzy Sets	Tập mờ
Frame	Khung hình
Incomplete	Không đầy đủ
Information Retrieval	Hệ thống tìm kiếm thông tin hỗ trợ
KD - tree	short for k - dimensional tree
Keypoint	Điểm quan trọng
Keypoint localization	Định vị điểm đặc trưng
Keypoint descriptor	Mô tả các điểm đặc trưng
L.A. Zadeh	Lotfi Aliasker Zadeh
Levenshtein distance	Khoảng cách Levenshtein
Lia, shot	Đoạn cơ sở
Match	Đối sánh
Necessity	Tất yếu
Optical flow	Dòng quang học
Orientation assignment	Xác định hướng
Panning	Chỉ sự quay trái, phải của máy quay
Possibility	Khả năng

Process	Tiến trình
RGB	Red, Green, Blue
Scale	Tỉ lệ
Scale - Space extrema detection	Phát hiện các điểm cực trị
Scene	Cảnh
Sequence	Chuỗi video, dãy video
SIFT	Scale Invariant Feature Transform
Spell - checker method	Phương pháp kiểm tra chính tả
Task	Tác vụ
TFN	Triangular Fuzzy Number
uncertainty	Không chắc chắn
unexact	Không chính xác
Weighted tuples	Bộ có trọng số

DANH MỤC BẢNG VÀ HÌNH VẼ

Hình 1.1: Biểu diễn các miền của một tập mờ.....	6
Hình 1.2: Biểu diễn các hàm liên thuộc của phép lấy phần bù.....	8
Hình 1.3: Các tập mờ hình tam giác	9
Hình 1.4: Tập mờ Singleton	9
Hình 1.5: Tập mờ L (phải)	9
Hình 1.6: Tập mờ Gamma tuyến tính	10
Hình 1.7: Tập mờ hình thang	10
Hình 2.1: Biểu diễn số mờ tam giác	24
Hình 2.2: Minh họa các bước chính trong giải thuật SIFT	26
Hình 2.3: Quá trình tính không gian đo (L) và hàm sai khác D.....	28
Hình 2.4: Quá trình tìm điểm cực trị trong các hàm sai khác DoG	29
Hình 2.5: Mô phỏng công thức mở rộng của Taylor cho hàm DoG	30
Hình 2.6: Minh họa các bước của quá trình lựa chọn các điểm keypoints	32
Hình 2.7: Tính độ lớn và hướng của Gradient.....	34
Hình 2.8: Tạo bộ mô tả cục bộ	35
Hình 2.9: Mô hình dữ liệu video	37
Hình 2.10: Sơ đồ khối thuật toán phát hiện chuyển cảnh	39
Hình 3.1: Mô hình bài toán tìm kiếm video trong CSDL	46
Hình 3.2: Quy trình tìm kiếm video bằng hình ảnh trong CSDL.....	47
Hình 3.3: Mô phỏng các điểm keypoint cần lưu trữ	48
Hình 3.4: Biểu diễn các tình huống trong trường hợp thông thường.....	50
Hình 3.5: Biểu diễn các tình huống trong trường hợp xấu	51
Hình 3.6: Sơ đồ khối đối sánh 2 số mờ tam giác	55
Hình 3.7: Giao diện chương trình phân đoạn và mờ hóa video	56
Hình 3.8: Giao diện chương trình Smart Cutter for DV and DVB	57
Hình 3.9: Mô hình lập trình socket TCP giữa 2 tiến trình client và server.....	59
Hình 3.10: Giao diện chương trình tìm kiếm video	61
Bảng dữ liệu thông tin Video.....	53
Bảng dữ liệu Video shot.....	54
Bảng kết quả đánh giá hiệu năng tìm kiếm.....	62

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC TỪ TIẾNG ANH VÀ VIẾT TẮT	iii
DANH MỤC BẢNG VÀ HÌNH VẼ	v
MỤC LỤC	vi
MỞ ĐẦU	1
1. Đặt vấn đề.....	1
2. Đối tượng và phạm vi nghiên cứu.....	2
3. Hướng nghiên cứu của đề tài.....	2
4. Những nội dung nghiên cứu chính.....	3
5. Phương pháp nghiên cứu	3
6. Ý nghĩa khoa học của đề tài.....	4
CHƯƠNG 1: TỔNG QUAN VỀ LÝ THUYẾT MỜ CSDL MỜ VÀ ỨNG DỤNG	5
1.1. Tổng quan về lý thuyết mờ.....	5
1.1.1. Khái niệm tập mờ	5
1.1.2. Một số khái niệm của tập mờ	5
1.1.3. Các phép toán trên tập mờ.....	6
1.1.4. Các kiểu hàm thuộc.....	8
1.1.5. Các loại số mờ.....	10
1.2. Cơ sở dữ liệu mờ, phương thức biểu diễn và ứng dụng.....	11
1.2.1. Định nghĩa	12
1.2.2. Biểu diễn thuộc tính trong quan hệ mờ.....	12
1.2.3. So sánh các giá trị thuộc tính mờ.....	13
1.2.4. Các phép toán tập hợp trên quan hệ mờ.....	14
1.2.5. Các phép toán quan hệ mờ	15
1.2.6. Phụ thuộc hàm mờ.....	16
1.2.7. Các mô hình cơ sở dữ liệu mờ.....	16
a. Mô hình tập con mờ.....	16

b. Mô hình dựa trên quan hệ tương tự.....	16
c. Mô hình dựa trên phân bố khả năng.....	17
d. Mô hình dựa trên phân bố khả năng mở rộng.....	18
e. Mô hình CSDL mờ dựa trên tổ hợp các mô hình trên	18
f. Mô hình theo cách tiếp cận đại số gia tử.....	18
1.3. Tổng quan về phương pháp tìm kiếm mờ.....	19
1.3.1. Khoảng cách Levenshtein.....	19
1.3.2. Khoảng cách Damerau - Levenshtein.....	20
1.3.3. Thuật toán Bitap với những thay đổi của Wu và Manber	20
1.3.4. Phương pháp kiểm tra chính tả.....	21
1.3.5. Phương pháp N-gram.....	21
1.3.6. Cây - BK (BK - trees).....	22
1.4. Kết luận chương	22
CHƯƠNG 2: THUẬT TOÁN TÌM KIẾM TRONG CSDL MỜ.....	23
2.1. Số mờ tam giác	23
2.1.1. Các phép toán trên số mờ tam giác	24
2.1.2. Quan hệ so sánh trên số mờ tam giác	24
2.2. Tổ chức dữ liệu cho thuật toán tìm kiếm dữ liệu mờ.....	25
2.2.1. Đối sánh ảnh dựa trên đặc trưng SIFT	25
a. Phát hiện các điểm cực trị.....	26
b. Định vị các điểm đặc trưng.....	30
c. Xác định hướng cho các điểm đặc trưng.....	33
d. Mô tả các điểm đặc trưng.....	34
e. Đối sánh đặc trưng SIFT	35
2.2.2. Phân đoạn dữ liệu video.....	37
a. Mô hình dữ liệu video	37
b. Một số kỹ thuật phân đoạn video	38
c. Kỹ thuật phân đoạn video bằng đối sánh đặc trưng SIFT.....	38
d. Tóm tắt video	40

2.3. Cải tiến thuật toán tìm kiếm dữ liệu mờ.....	41
2.3.1. Thuật toán tìm kiếm	41
2.3.2. Thuật toán tìm kiếm dữ liệu mờ.....	42
2.3.3. Đánh giá thuật toán tìm kiếm.....	43
2.4. Kết luận chương	44
CHƯƠNG 3: CÀI ĐẶT THUẬT TOÁN TÌM KIẾM	45
3.1. Phân tích một số nhu cầu tìm kiếm video trong CSDL.....	45
3.2. Thiết kế mô hình bài toán tìm kiếm trong CSDL	46
3.2.1. Mô hình bài toán tìm kiếm video.....	46
3.2.2. Quy trình tìm kiếm video theo hình ảnh	47
3.2.3. Quy trình sử dụng logic mờ	48
3.2.4. Mở rộng csdl quan hệ thành csdl mờ dựa trên phân bố khả năng	49
a. Bộ có trọng số trong quan hệ mờ	49
b. Biểu diễn dữ liệu mờ bằng phân bố khả năng.....	50
c. Môi quan hệ với dạng chuẩn 1NF.....	52
d. Độ đo khả năng và độ đo cần thiết	52
3.3. Cài đặt thuật toán tìm kiếm mờ và một số giao diện chính.....	53
3.3.1. Lựa chọn công cụ.....	53
3.3.2. Thiết kế cơ sở dữ liệu	53
3.3.3. Cài đặt thuật toán tìm kiếm và một số giao diện chính.....	55
a. Kỹ thuật đối sánh hai số mờ tam giác	55
b. Cài đặt thuật toán phân đoạn video	56
c. Cài đặt thuật toán trích chọn frame đại diện và mờ hóa video	57
d. Kỹ thuật khởi tạo các tiến trình và trao đổi dữ liệu.....	59
e. Cài đặt thuật toán tìm kiếm dữ liệu mờ	60
3.3.4. Thử nghiệm và đánh giá.....	62
3.4. Kết luận chương	62
KẾT LUẬN	63
TÀI LIỆU THAM KHẢO	65