

**BỘ GIÁO DỤC VÀ  
ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA  
HỌC VÀ CÔNG NGHỆ  
VIỆT NAM**

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**  
.....

**NGUYỄN THỊ LAN PHƯƠNG**

**MỘT SỐ KỸ THUẬT NÂNG CAO HIỆU QUẢ TRA CỨU ẢNH  
THEO NỘI DUNG DỰA TRÊN ĐỘ ĐO KHOẢNG CÁCH THÍCH  
NGHI VÀ PHÂN CỤM PHỒ**

**Ngành: Khoa học máy tính  
Mã số: 9 48 01 01**

**TÓM TẮT LUẬN ÁN TIẾN SĨ MÁY TÍNH**

**Hà Nội – 2023**

**Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ  
- Viện Hàn lâm Khoa học và Công nghệ Việt Nam**

**Người hướng dẫn khoa học 1: PGS.TS. Ngô Quốc Tạo**

**Người hướng dẫn khoa học 2: TS. Nguyễn Ngọc Cương**

**Phản biện 1: PGS.TS. Đỗ Trung Tuấn**

**Phản biện 2: PGS.TS. Nguyễn Đức Dũng**

**Phản biện 3: TS. Nguyễn Văn Căn**

Luận án được bảo vệ trước Hội đồng chấm luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ, ngày tháng năm 2023

**Có thể tìm hiểu luận án tại:**

- Thư viện Học viện Khoa học và Công nghệ
- Thư viện Quốc gia Việt Nam

## PHẦN MỞ ĐẦU

### 1. Tính cấp thiết của luận án

Trong thập kỷ qua, chúng ta đã chứng kiến sự tăng trưởng liên tục của số lượng ảnh kỹ thuật số được chụp, lưu trữ và chia sẻ mỗi ngày. Ước tính số lượng ảnh kỹ thuật số được chụp năm 2021 là hơn 5 nghìn tỷ. Khoảng 85% trong số đó là chụp bằng điện thoại di động. Một phần lớn trong số chúng có sẵn trên Internet thông qua các trang web, thư viện ảnh (Flickr và Shutterstock), và các phương tiện truyền thông xã hội khác nhau Facebook, Instagram.... Phần lớn các cơ sở dữ liệu ảnh này, không được sắp xếp cũng không đính kèm siêu dữ liệu và thẻ. Ngoài ra, cơ sở dữ liệu ảnh phổ biến trong các lĩnh vực ứng dụng như phòng chống tội phạm, y học, kiến trúc, viễn thám,... Các kỹ thuật thu truyền và lưu trữ ảnh ngày càng phát triển đã cho phép xây dựng các cơ sở dữ liệu ảnh khổng lồ. Tra cứu ảnh dựa vào nội dung (CBIR) giải quyết bài toán quản lý thư viện ảnh, phân loại ảnh, nhận dạng đối tượng trong ảnh, tra cứu hình ảnh trên mạng và nhiều ứng dụng khác liên quan đến xử lý ảnh và thị giác máy tính. Do vậy, việc tra cứu nhanh chóng và chính xác một bức ảnh mong muốn trong một cơ sở dữ liệu ảnh số lớn và đa dạng là một nhiệm vụ hết sức khó khăn, đầy thách thức trong lĩnh vực thị giác máy tính hiện nay.

### 2. Mục tiêu của luận án

*Mục tiêu chung của luận án:*

Đề xuất được phương pháp tra cứu ảnh cho nâng cao độ chính xác tra cứu.

*Mục tiêu cụ thể của luận án:*

- Cải tiến phương pháp tra cứu ảnh bằng phương pháp ODLDA thông qua tìm một phép đo khoảng cách tối ưu, mà giảm khoảng cách giữa các cặp ảnh có độ tương tự cao và tối đa hóa khoảng cách giữa các cặp ảnh có độ tương tự thấp.

- Đề xuất phương pháp tra cứu ảnh dựa trên lý thuyết cắt đồ thị, mà không phải tính ma trận Laplacian, các giá trị riêng và các véc tơ riêng.

### 3. Đối tượng nghiên cứu

Đối tượng nghiên cứu của luận án là tra cứu ảnh dựa trên nội dung bằng cách kết hợp khoảng cách tối ưu và phân tích phân biệt tuyến tính, tiến hành thực nghiệm trên tập cơ sở dữ liệu tập ảnh Corel (1 0.800 ảnh), phân hoạch đồ thị với cơ sở dữ liệu ảnh SIMPLicity (1.000 ảnh với 10 chủ đề. Mỗi ảnh có kích thước 256×384 hoặc 384×256).

#### 4. Phương pháp nghiên cứu của luận án

Phương pháp nghiên cứu của luận án là nghiên cứu lý thuyết và nghiên cứu thực nghiệm. Về nghiên cứu lý thuyết: giới thiệu về tra cứu ảnh dựa vào nội dung, một số nghiên cứu ảnh dựa vào nội dung, trích rút đặc trưng, thông tin không gian, đo khoảng cách, phân cụm, giảm khoảng cách ngữ nghĩa, phân tích phân biệt tuyến tính, đánh giá hiệu năng.

#### 5. Bố cục của luận án

Luận án này được bố cục thành ba chương:

Chương 1: Tổng quan về tra cứu ảnh dựa trên nội dung.

Chương 2: Nâng cao hiệu quả của việc tra cứu ảnh dựa trên nội dung bằng cách kết hợp tối ưu khoảng cách và phân tích phân biệt tuyến tính.

Chương 3: Cải thiện hiệu quả của tra cứu ảnh dựa trên nội dung sử dụng phân hoạch đồ thị

Cuối cùng, luận án đưa ra một số kết luận và định hướng nghiên cứu trong tương lai.

#### 6. Kết quả và tính mới của luận án

Đóng góp vào hướng nghiên cứu, luận án đưa ra được những đóng góp sau:

- (1) Luận án nâng cao độ chính xác tra cứu ảnh thông qua việc xây dựng cơ sở dữ liệu véc tơ đặc trưng với mạng học sâu CNN AlexNet.
- (2) Trong quá trình học độ đo tương tự, luận án xem xét cả tập liên quan và tập không liên quan và sử dụng phương pháp học phân tích phân biệt tuyến tính LDA để tiến hành điều chỉnh hàm trọng số của hàm khoảng cách.
- (3) Đề xuất phương pháp tra cứu ảnh hiệu quả sử dụng phân hoạch đồ thị (An efficient image retrieval method using a graph clustering-MGC) mà khai thác đầy đủ thông tin độ tương tự của tập ảnh. Kết quả thực nghiệm của luận án trên cơ sở dữ liệu đặc trưng gồm 1.000 ảnh đã chỉ ra rằng phương pháp được đề xuất **MGC** cung cấp một độ chính xác cao hơn so với các phương pháp khác.

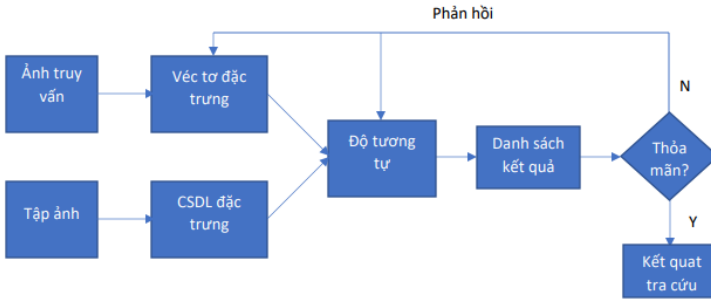
### Chương 1. TỔNG QUAN VỀ TRA CỨU ẢNH DỰA VÀO NỘI DUNG

Chương này giới thiệu cơ bản về tra cứu ảnh dựa vào nội dung bao gồm: xem xét sự phát triển của kỹ thuật tra cứu ảnh dựa vào nội dung, cách mô tả nội dung trực quan, độ đo khoảng cách giữa các nội dung trực quan,

lược đồ chỉ mục, tạo truy vấn, cơ chế phản hồi liên quan. Bên cạnh đó, chương này cũng trình bày về đánh giá hiệu năng hệ thống. Cuối cùng, chương này đưa ra một số kết luận và định hướng cho nghiên cứu.

### 1.1. Giới thiệu

Tra cứu ảnh dựa trên nội dung (CBIR), sử dụng nội dung trực quan của ảnh như màu sắc, hình dạng, kết cấu và bố cục không gian để biểu diễn và lập chỉ mục cho hình ảnh. Trong các hệ thống CBIR điển hình (Hình 1.1),



Hình 1.1. Sơ đồ tra cứu ảnh dựa vào nội dung

### 1.2. Mô tả nội dung ảnh

Nhìn chung, nội dung ảnh có thể bao gồm nội dung trực quan của ảnh và nội dung ngữ nghĩa của ảnh. Nội dung trực quan có thể rất chung hoặc theo miền cụ thể. Nội dung ảnh chung bao gồm màu sắc, kết cấu, hình dạng, và quan hệ không gian,..... Màu sắc

#### 1.2.1. Không gian màu

Không gian RGB là không gian màu được sử dụng rộng rãi để hiển thị hình ảnh. Nó bao gồm ba thành phần là màu đỏ, màu xanh lá cây và màu xanh lam.

#### 1.2.2. Mô men màu

Mô men màu đã được sử dụng thành công trong hệ thống tra cứu ảnh [1,2] (như QBIC), đặc biệt là khi ảnh chỉ chứa đối tượng. Mô men màu bậc nhất (trung bình), bậc hai (phương sai) và bậc 3 (độ lệch) đã được chứng minh là có hiệu quả và hiệu quả trong việc biểu diễn phân bố màu của ảnh. .

### 1.2.3. *Biểu đồ màu*

Biểu đồ màu là sự biểu diễn hiệu quả nội dung màu của ảnh nếu màu là duy nhất so với phần còn lại của tập dữ liệu.

### 1.2.4. *Biểu đồ màu tương quan*

Biểu đồ tương quan màu được đề xuất để mô tả không chỉ sự phân bố màu sắc của các pixel, mà còn cả mối tương quan trong không gian của các cặp màu [17]. Chiều thứ nhất và thứ hai của biểu đồ ba chiều là màu của bất kỳ cặp pixel nào và chiều thứ ba là khoảng cách không gian của chúng.

### 1.2.5. *Đặc trưng màu*

Màu sắc không chỉ phản chiếu chất liệu bề mặt mà còn thay đổi đáng kể theo sự thay đổi của độ chiếu sáng, hướng của bề mặt và hình dạng quan sát của máy ảnh [19,20]. Sự thay đổi này phải được tính đến. Tuy nhiên, sự bất biến đối với các yếu tố môi trường này không được xem xét trong hầu hết các màu sắc được giới thiệu ở trên.

Gần đây, biểu diễn bất biến màu đã được giới thiệu trong tra cứu ảnh dựa trên nội dung. Một tập hợp các bất biến màu cho tra cứu đối tượng được suy diễn dựa trên mô hình phản xạ đối tượng của Schafer. Biểu diễn bất biến phản xạ, hình dạng và độ chiếu sáng dựa trên véc tơ tỉ lệ xanh lam ( $r/b$ ,  $g/b$ , 1) được đưa ra. Trong [20], đặc trưng bất biến hình học bề mặt được cung cấp.

Mô men màu bất biến này được áp dụng để tra cứu ảnh, có thể mang lại khả năng chiếu sáng, và biểu diễn hình học độc lập với nội dung màu của hình ảnh, nhưng cũng có thể dẫn đến mất một số khả năng phân biệt giữa các hình ảnh.

### 1.2.6. *Đặc trưng kết cấu.*

### 1.2.7. *Đặc trưng Tamura*

Tamura bao gồm độ thô, độ tương phản, tính định hướng, độ đều và độ nhám, được thiết kế phù hợp với các nghiên cứu tâm lý về nhận thức của con người về kết cấu. Ba thành phần đầu tiên của Tamura đã được sử dụng trong một số hệ thống tra cứu ảnh nổi tiếng ban đầu, chẳng hạn như QBIC và Photobook.

### 1.2.8. *Độ thô*

Độ thô là thước đo độ chi tiết của kết cấu.

**1.2.9. Độ tương phản****1.2.10. Mô hình tự hồi quy đồng thời****1.2.11. Bộ lọc Gabor**

Bộ lọc Gabor đã được sử dụng rộng rãi để trích rút đặc trưng của hình ảnh, đặc biệt là các đặc trưng về kết cấu.

**1.2.12. Biến đổi Wavelet****1.2.13. Đặc trưng hình dạng****1.2.14. Mô men bất biến**

Mô men bất biến được gọi là ‘invariant moment’ là tập hợp các đặc trưng số học của hình ảnh được tính toán dựa trên các giá trị cường độ của điểm ảnh trong hình ảnh. Mục đích của việc sử dụng mô men bất biến là để tạo ra các đặc trưng có tính chất không thay đổi khi ảnh bị thay đổi bởi các biến đổi hình học như quay, phóng to, thu nhỏ hoặc lật đối xứng, điều này giúp cho việc nhận dạng và phân loại đối tượng trở nên ổn định hơn trong các tình huống khác nhau.

**1.2.15. Góc quay**

Góc quay thể hiện mức độ xoay của hình ảnh quanh một trục tương ứng. Trong không gian hai chiều, góc quay được đo bằng độ và thường được tính theo chiều kim đồng hồ. Trong xử lý ảnh, để biến đổi xoay thường sử dụng biến đổi hình học như ma trận xoay. Ma trận xoay  $2 \times 2$  và góc quay được tính theo radian. Ma trận xoay áp dụng lên các điểm ảnh trong hình ảnh để thực hiện biến đổi xoay. Biến đổi xoay sử dụng trong việc tạo ra các phiên bản xoay của ảnh để tạo ra dữ liệu đào tạo đa dạng hơn trong mô hình học máy.

**1.2.16. Mô tả Fourier**

Biến Fourier là một công cụ tốt trong xử lý ảnh giúp phân tích và xử lý tín hiệu ảnh dựa trên phổ tần số của chúng.

**1.2.17. Tính tuần hoàn, độ lệch tâm và hướng trục chính**

Tính tuần hoàn được tính là :

$$\alpha = \frac{4\pi S}{P^2} \quad (1.20)$$

Trong đó, S là kích thước và P là chu vi của một vật thể. Giá trị này tương ứng với một đường tròn hoàn hảo.

Hướng trục chính có thể được xác định là hướng của ký hiệu riêng lớn nhất của ma trận hiệp phương sai bậc hai của một vùng hoặc một đối

trương. Độ lệch tâm có thể được định nghĩa là tỷ số giữa giá trị riêng nhỏ nhất và giá trị riêng lớn nhất.

### **1.2.18. Thông tin không gian**

## **1.3. Các kỹ thuật tương tự và các lược đồ lập chỉ mục**

### **1.3.1. Khoảng cách Minkowski**

### **1.3.2. Khoảng cách toàn phương**

### **1.3.3. Khoảng cách Mahalanobis**

### **1.3.4. Lập chỉ mục**

## **1.4. Tương tác người dùng**

Đối với tra cứu hình ảnh dựa trên nội dung, tương tác của người dùng với hệ thống tra cứu là rất quan trọng vì nó có thể sửa đổi linh hoạt các truy vấn bằng cách để người dùng tham gia vào quá trình tra cứu. Giao diện người dùng trong hệ thống tra cứu hình ảnh bao gồm phần tạo truy vấn và phần trình bày kết quả.

### **1.4.1. Kỹ thuật truy vấn bởi phác thảo**

### **1.4.2. Phản hồi liên quan**

### **1.4.3. Đánh giá hiệu năng**

## **1.5. Giảm khoảng cách ngữ nghĩa**

### **1.5.1. Khái niệm**

Khoảng cách ngữ nghĩa là một trong những ví dụ điển hình trong tra cứu ảnh dựa vào nội dung. Khoảng cách ngữ nghĩa là khoảng cách đề cập đến mức độ tương đồng hoặc sự giống nhau (khoảng cách) giữa nhận thức của con người và sự hiểu biết có được từ các thuật toán máy tính về cùng một ảnh. Khoảng cách này có ảnh hưởng trực tiếp đến việc đánh giá các ảnh là tương tự bởi các thuật toán. Sự tương tự về ảnh được xác định bởi một người quan sát trong ngữ cảnh cụ thể ở cấp độ ngữ nghĩa cao.

### **1.5.2. Một số nghiên cứu theo hướng tiếp cận học có giám sát.**

### **1.5.3. Một số nghiên cứu theo hướng tiếp cận học không giám sát**

## **1.6. Phân tích phân biệt tuyến tính**

### **1.6.1. Phân tích phân biệt tuyến tính cho bài toán với hai lớp**

#### **1.6.1.1 Ý tưởng cơ bản.**

#### **1.6.1.2. Xây dựng hàm mục tiêu**

## **1.7 Kết luận chương 1**

Với dữ liệu ảnh lớn như hiện nay và lượng ảnh tăng lên theo từng giờ, từng ngày, việc nghiên cứu các phương pháp CBIR hiệu quả cực kỳ cần



thiết. Và đối với hệ thống CBIR việc tăng độ chính xác tra cứu ảnh và tăng tốc độ tra cứu ảnh là hai việc cần làm đầu tiên và cần thiết. Để làm được hai việc này thì hệ thống CBIR phải tập trung vào hai giai đoạn quan trọng nhất là trích rút đặc trưng và tính độ tương tự.

## **Chương 2: NÂNG CAO HIỆU QUẢ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG BẰNG CÁCH KẾT HỢP KHOẢNG CÁCH TỐI ƯU VÀ PHÂN TÍCH PHÂN BIỆT TUYẾN TÍNH**

Việc tra cứu ảnh dựa trên nội dung được thực hiện bằng cách so sánh sự tương tự giữa biểu diễn ảnh truy vấn và từng biểu diễn ảnh trong cơ sở dữ liệu. Do đó, biểu diễn ảnh và độ đo tương tự là hai phần cốt lõi của tra cứu ảnh dựa trên nội dung. Trong tra cứu ảnh với phản hồi liên quan, tính toán khoảng cách và phân lớp có một ảnh hưởng lớn lên độ chính xác tra cứu ảnh. Trong chương này, luận án trình bày phương pháp tra cứu ảnh đề xuất, gọi là ODLDA (Image Retrieval using the optimal distance and linear discriminant analysis). Phương pháp đề xuất có thể khai thác phản hồi của người dùng từ tập các ảnh liên quan và không liên quan, mà sử dụng phân tích phân biệt tuyến tính để tìm một chiều tuyến tính với một độ đo tương tự cải tiến. Các kết quả thực nghiệm thực hiện trên hai tập dữ liệu tiêu chuẩn đã thấy sự tiến bộ của phương pháp đề xuất. Phương pháp đề xuất có thể khai thác hiệu quả phản hồi của người dùng từ tập hợp ảnh không liên quan, sử dụng phân tích phân biệt tuyến tính để tìm một phép chiếu tuyến tính với một số đo tương tự được cải thiện.

### **2.1. Giới thiệu**

Trong một hệ thống CBIR tiêu biểu, các đặc trưng trực quan mức thấp bao gồm màu, kết cấu, và hình dạng, mà được trích rút tự động và được biểu diễn thành các véc tơ đặc trưng.

Ý tưởng của học độ đo tương tự là để tìm một độ đo khoảng cách tối ưu mà cực tiểu khoảng cách giữa các cặp ảnh tương tự và cực đại khoảng cách giữa các cặp ảnh không tương tự. Sau đó, độ đo khoảng cách tối ưu này được sử dụng để phân hạng lại toàn bộ tập ảnh và trả lại các kết quả tốt hơn. Trong luận án, nghiên cứu sinh đề xuất một kỹ thuật tra cứu ảnh hiệu quả (ODLDA). Phương pháp đề xuất chính xác hơn một số phương pháp đã có bởi vì biểu diễn đặc trưng là có ngữ nghĩa cao hơn và các độ đo tương tự

được học là phù hợp với dữ liệu hơn. Bằng thực nghiệm với hai cơ sở dữ liệu tiêu chuẩn, độ chính xác của phương pháp được đề xuất được chỉ ra.

## 2.2. Nghiên cứu liên quan

Học độ đo tương tự trong tra cứu ảnh dựa vào nội dung đã nhận được sự chú ý của cộng đồng nghiên cứu. Trong tra cứu ảnh với phản hồi liên quan, dữ liệu đầu vào của các thuật toán học độ đo tương tự thường được chia thành hai nhóm: nhóm thứ nhất gồm các cặp ảnh tương tự; và nhóm thứ hai gồm các cặp ảnh tương tự và các cặp ảnh không tương tự.

Trong phương pháp MCML.

Ý tưởng của phương pháp LMNN.

Ý tưởng trong phương pháp của Xing.

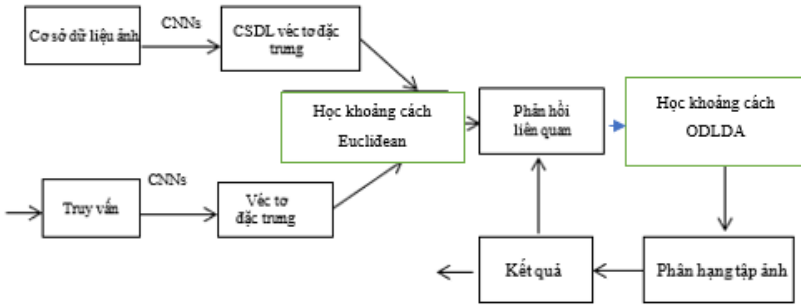
Ý tưởng của phương pháp RCA.

Từ phân tích giới hạn của các nghiên cứu liên quan ở trên, luận án đề xuất một phương pháp tra cứu ảnh cải tiến. Cải tiến hàm khoảng cách dựa trên cực đại tỉ số giữa tổng khoảng cách của các cặp ảnh không tương tự và tổng khoảng cách của các cặp ảnh tương tự. Ở đây, NCS xét cả tập các ảnh tương tự và không tương tự để tìm ma trận trọng số và cải tiến độ chính xác của tra cứu ảnh.

## 2.3. Đề xuất phương pháp phân hạng lại ảnh

Trong phần này, luận án trình bày ngắn gọn phương pháp đề xuất. Đầu tiên, phương pháp đề xuất việc xây dựng các đặc trưng sâu để biểu diễn các ảnh. Tiếp theo, trên tập kết quả của pha tra cứu khởi tạo mà sử dụng các đặc trưng sâu, người dùng đánh dấu các ảnh liên quan đến ảnh truy vấn để thu được tập ảnh phản hồi. Tập phản hồi thu được này bao gồm các ảnh liên quan và không liên quan đến ảnh truy vấn. Dựa trên tập ảnh liên quan, phương pháp đề xuất huấn luyện mô hình để tìm phép chiếu tuyến tính. Phép chiếu tuyến tính này thỏa mãn điều kiện mà phương sai giữa các mẫu trong cùng tập liên quan được cực tiểu trong khi cực đại phương sai giữa các mẫu liên quan với các mẫu không liên quan. Bên cạnh đó, phương pháp đề xuất cũng xây dựng một độ đo tương tự Mahalanobis bằng việc tìm ma trận tối ưu  $M$  trong công thức độ đo tương tự cải tiến.

### 2.3.1. Sơ đồ của phương pháp đề xuất



Hình II. 1. Sơ đồ của phương pháp đề xuất ODLDA

### 2.3.2. Tra cứu ảnh sử dụng học sâu

Trong những năm gần đây, mạng CNN đã cho các kết quả tốt trong lĩnh vực thị giác máy tính như phân lớp ảnh, nhận dạng đối tượng, phân đoạn ngữ nghĩa. Trên cơ sở đó, đã có những nghiên cứu về tra cứu ảnh dựa vào nội dung sử dụng CNN và đã thu được các kết quả khả quan.

Giả sử NCS có hai ảnh trong cơ sở dữ liệu  $I_1$  và  $I_2$ , các đặc trưng sâu được trích rút sử dụng mô hình CNN được tiền huấn luyện trên tập dữ liệu ImageNet. Đặc trưng sâu của hai ảnh  $I_1$  và  $I_2$  được biểu thị bởi  $x_1$  và  $x_2$ . Độ đo tương tự được sử dụng để so sánh hai đặc trưng này là  $L_2$ :

$$\begin{aligned}
 L_2\_Similarity(x_i, x_j) &= \|x_i - x_j\|_2 \\
 &= \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (2.1)
 \end{aligned}$$

Công thức (2.1) chỉ ra độ tương tự giữa các ảnh  $I_i$  và  $I_j$ , giá trị độ tương tự là lớn hơn cho các ảnh  $I_i$  và  $I_j$  giống nhau hơn.

Độ đo tương tự sử dụng cách tiếp cận 2) để so sánh hai véc tơ đặc trưng của ảnh được tính bởi công thức  $L_T$ :

$$L_T\_similarity(x_i, x_j) = \|x_i - x_j\|_T$$

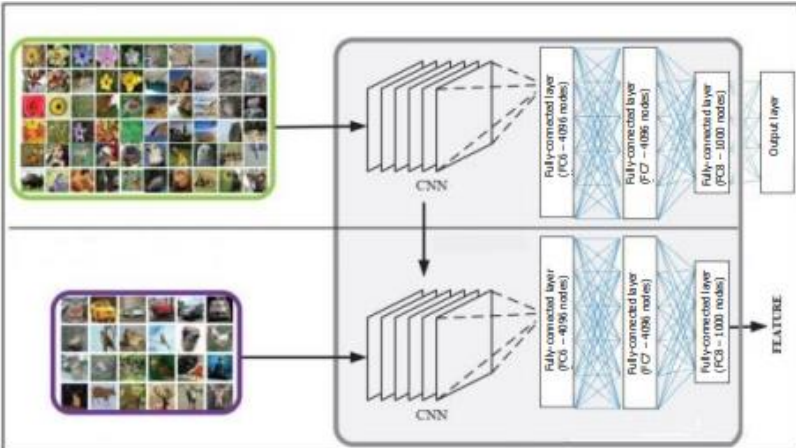
$$= \sqrt{(x_i - x_j)^T T(x_i - x_j)} \quad (2.2)$$

Với một ma trận, thu được từ việc học chỉ số tương tự thỏa mãn điều kiện là ma trận xác định dương, vì chỉ số tương tự phải dương và chỉ số tương tự có giá trị nhỏ nhất khi

$$x_i = x_j$$

Với một ma trận,  $T$  thu được từ học độ đo tương tự mà thỏa mãn điều kiện  $T$  là một ma trận xác định dương, bởi vì độ đo tương tự phải là dương, và độ đo tương tự có giá trị nhỏ nhất khi  $x_1 = x_2$ .

Độ đo tương tự ở đây là như trong cách tiếp cận 1) khi ma trận  $T$  là một ma trận đơn vị  $T = I$ . Nói cách khác, nó là một trường hợp đặc biệt khi xét tương quan giữa các thành phần đặc trưng trong cách tiếp cận 1). Hơn nữa, mỗi thành phần đặc trưng có một sự tương tự khác nhau, vậy nó thường là độ đo tương tự trong cách tiếp cận 2) để thu được hiệu quả cao hơn.



Hình II. 2. Kiến trúc học biểu diễn dựa vào mô hình CNN được tiền huấn luyện.

## 2.4. Độ đo khoảng cách cải tiến .

### 2.5. Thuật toán tra cứu ảnh

Thuật toán 1.1, gọi là ODLDA, là thuật toán tra cứu ảnh dựa vào phân tích phân biệt tuyến tính và khoảng cách tối ưu.

---

#### Algorithm 1.1.ODLDA

---

##### Input:

Image set :  $DB$

Initialization query image:  $Q$

Returned image number for each iteration :  $N$

##### Output:

Result:  $R$

1.  $S \leftarrow \text{IRL}\langle DB, M \rangle$ ;

2.  $S_q \leftarrow \text{IRL}\langle Q, M \rangle$ ;

3.  $\text{Result}_{\text{initial}}(Q) \leftarrow \text{Retrieval}_{\text{Initial}}(S_q, S, N)$

4.  $R \leftarrow \text{Result}_{\text{initial}}(Q)$ ;

##### 5. Repeat

5.1.  $\langle F_{\text{feature}}, F_{\text{label}}^+, F_{\text{label}}^- \rangle \leftarrow \text{Feedback}(R)$ ; *relevant feedback*

5.2.  $W = \text{LDA}(F_{\text{feature}}, F_{\text{label}}^+, F_{\text{label}}^-)$ ; *Find the optimal transformation W*

5.3.  $W_o = W^T W$ ; *The optimal weight of the Mahalanobis distance function*

5.4.  $R \leftarrow \text{Ranking}(S, W_o, N)$ ; *Rerank the set of images according to the Mahalanobis distance function with the optimal weight.*

**Until** (User stops responding);

**Kmeans**( $f$ )

6. **Return**  $R$  ;

---

## 2.6. Kết quả thực nghiệm

### 2.6.1. Môi trường thực nghiệm

1) Tập dữ liệu ảnh Corel:

2) Tập tin cậy nền (Ground truth) cho đánh giá độ chính xác của CBIR: tập tin cậy nền được sử dụng để đánh giá độ chính xác của hệ thống CBIR, tức là, các ảnh liên quan và không liên quan được biết trước ở trong tập tin cậy nền này. Theo đó, hệ thống tra cứu ảnh xem xét các ảnh mà liên quan đến ảnh truy vấn là các ảnh có cùng chủ đề. Tập này gồm ba cột (tiêu đề : Query Image ID, Image ID, and Relation) và bao gồm 1,981,320 dòng.

3) Tập ảnh SIMPLIcity: Để minh chứng hiệu năng của phương pháp đề xuất.



Hình II. 3. Một số mẫu trong thư viện ảnh Corel.



Hình II. 4. Một số mẫu trong tập SIMPLIcity.

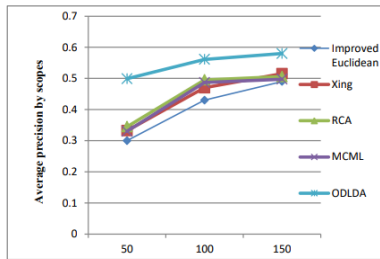
### 2.6.2. Đánh giá thực nghiệm

Độ chính xác trung bình của các phương pháp được chỉ ra trên Bảng II.1. Trong bảng này, luận án thấy rằng phương pháp sử dụng độ đo khoảng cách Euclidean gốc có độ chính xác thấp nhất. Ba phương pháp Xing, RCA, và MCML có độ chính xác trung tự. Phương pháp đề xuất có độ chính xác cao nhất.

Phương pháp	Độ chính xác trung bình theo các phạm vi (scope)
-------------	--

	50	100	150
<b>Euclide</b>	0.2887	0.3065	0.3199
<b>Euclide cải tiến (Improved Euclidean)</b>	0.3135	0.42658	0.4846
<b>Xing</b>	0.3324	0.47658	0.5125
<b>RCA</b>	0.3424	0.48058	0.5015
<b>MCML</b>	0.3328	0.47958	0.4925
<b>ODLDA</b>	<b>0.4836</b>	<b>0.5065</b>	<b>0.5199</b>

*Bảng II. 1. So sánh độ chính xác trung bình của các phương pháp ở scope 50, 100 và 150 trên tập dữ liệu Corel.*



*Hình II. 5. So sánh độ chính xác trung bình của các phương pháp trên các scope 50, 100 và 150 trên tập SIMPLiCity.*

## 2.7. Kết luận chương 2

L luận án trình bày phương pháp ODLDA, một kỹ thuật tra cứu ảnh hiệu quả kỹ thuật cải thiện hiệu suất của hệ thống tra cứu ảnh đa điểm. ODLDA khai thác hiệu quả thông tin của người dùng thông qua tập mẫu có liên quan và không liên quan, thực hiện học phép chiếu tối ưu để tách các ảnh không liên quan và thu hẹp khoảng cách của các ảnh liên quan. Phương pháp được đề xuất tìm ma trận trọng số tối ưu của hàm khoảng cách Mahalanobis và sử dụng hàm khoảng cách cải tiến này để xếp hạng toàn bộ tập ảnh cơ sở dữ liệu và trả về tập kết quả cho người dùng. Kết quả thử

nghiệm trên hai cơ sở dữ liệu đã chứng minh rằng ODLDA cung cấp độ chính xác cao hơn nhiều so với phương pháp Euclid, Euclid, RCA và OASIS cải tiến.

Kết quả thực nghiệm trên cơ sở dữ liệu đặc trưng gồm 1000 ảnh đã chỉ ra rằng phương pháp được đề xuất cung cấp một độ chính xác cao hơn hẳn so với các phương pháp khác .

Một phần công trình nghiên cứu tại CT4.

### **Chương 3. CẢI THIỆN HIỆU QUẢ CỦA TRA CỨU ẢNH DỰA TRÊN NỘI DUNG SỬ DỤNG PHÂN HOẠCH ĐỒ THỊ**

Trong những năm gần đây, nhiều phương pháp tra cứu ảnh (CBIR) theo cách tiếp cận phản hồi có liên quan được thiết kế để thu hẹp khoảng trống ngữ nghĩa giữa các đặc trưng trực quan mức thấp và các khái niệm ngữ nghĩa mức cao cho nhiệm vụ tra cứu ảnh. Tuy nhiên, các phương pháp tra cứu ảnh hiện nay chỉ quan tâm đến độ tương tự giữa ảnh truy vấn và ảnh cơ sở dữ liệu mà chưa quan tâm đến độ tương tự giữa các ảnh trong tập ảnh đích. Trong luận án này Nghiên cứu sinh đề xuất một phương pháp tra cứu ảnh hiệu quả sử dụng phân hoạch đồ thị (*An efficient image retrieval method using a graph clustering-MGC*) mà khai thác đầy đủ thông tin độ tương tự của tập ảnh. Phần thực nghiệm trên cung cấp các kết quả thực nghiệm để minh chứng độ chính xác của phương pháp đề xuất.

#### **3.1. Nâng cao hiệu quả tra cứu ảnh dựa vào nội dung sử dụng phân hoạch đồ thị**

##### **3.1.1. Giới thiệu**

Trong xử lý ảnh, đồ thị và phân hoạch đồ thị là các khái niệm quan trọng được sử dụng để mô tả và phân tích các đặc điểm của hình ảnh để cải thiện nâng cao tra cứu ảnh dựa vào nội dung.

Tuy nhiên, các phương pháp tra cứu ảnh sử dụng phản hồi liên quan đề cập ở trên có hạn chế: chỉ quan tâm đến độ tương tự giữa ảnh truy vấn và ảnh cơ sở dữ liệu mà chưa quan tâm đến độ tương tự giữa các ảnh trong tập ảnh đích. Vậy, có thể nâng cao hiệu năng của hệ thống tra cứu ảnh theo cách tiếp cận phản hồi liên quan bằng cách khai thác thông tin tương tự giữa các ảnh trong tập ảnh đích không?

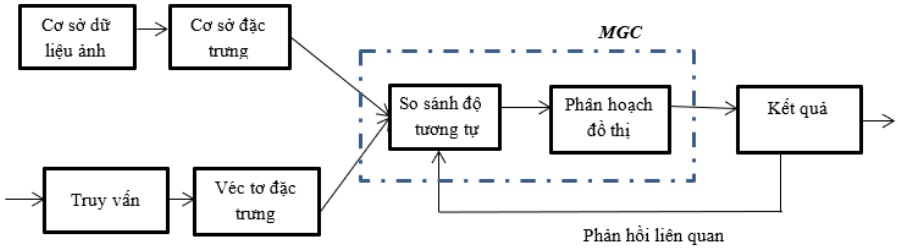


Đây là câu hỏi mà nghiên cứu sinh sẽ giải quyết trong nội dung “Nâng cao hiệu quả tra cứu ảnh dựa vào nội dung sử dụng phân hoạch đồ thị”.

### 3.1.2. Nghiên cứu liên quan:

### 3.1.3. Phương pháp đề xuất:

Phương pháp **MGC** được mô tả bởi lược đồ trên Hình III.1.



Hình III. 1. Sơ đồ của tra cứu ảnh sử dụng phân hoạch đồ thị

### 3.1.4. Phân cụm cắt tối thiểu lặp (Iterative Min Cut Clustering)

#### Iterative Min Cut Clustering

Phương pháp Iterative Min Cut Clustering (IMC) được đề xuất phân cùng một tập dữ liệu  $X = \{x_1, \dots, x_N\} \subset R^H$  thành  $C$  cụm bằng cách tối thiểu hóa hàm mục tiêu:

$$\sum_{i,j} w_{ij}, x_i \text{ và } x_j \text{ thuộc các cụm khác nhau} \quad (3.1)$$

với  $w_{ij}$  là độ tương đồng (trọng số cạnh) giữa  $x_i$  và  $x_j$ . Để việc tính toán cho thuận tiện, ta chuẩn hóa các điểm dữ liệu  $x_i$  ( $i \in \{1, \dots, N\}$ ) như sau:

$$x_i = \frac{x_i}{\max\{x_i[1], \dots, x_i[H]\}} \quad (3.2)$$

Độ tương tự  $w_{ij}$  được tính bằng:

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), & x_i \text{ và } x_j \text{ là các láng giềng} \\ 0 & \text{nếu ngược lại} \end{cases}$$

-(3.3)

Để giải quyết vấn đề (1), ta định nghĩa một feature  $q$  (là đại lượng vô hướng) cho mỗi điểm dữ liệu. Nếu 2 điểm dữ liệu thuộc cùng một cụm thì  $q$  của chúng sẽ có giá trị giống nhau và ngược lại. Có  $q_i$  đại diện cho feature của  $x_i$ ,  $q_i = q_j$  nếu  $x_i$  và  $x_j$  thuộc cùng một cụm và  $q_i \neq q_j$  nếu ngược lại.

Véc tơ  $q = [q_i] = [q_1, \dots, q_N]^T$  có thể được xem như một chiều được gán của tập dữ liệu  $X$ . (1) tương đương với:

$$Q = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (q_i - q_j)^2 \quad (3.4)$$

Dựa vào mối quan hệ giữa (4) và ma trận Laplacian:

$$q^T L q = \frac{1}{2} \sum_{i,j} w_{ij} (q_i - q_j)^2 \quad (3.5)$$

Để giải quyết vấn đề (3.4):

$$\frac{\partial Q}{\partial q_i} = 2 \sum_j (q_i - q_j) w_{ij} - 2j \sum_j (q_i - q_j) w_{ji} = 4 \sum_j (q_i - q_j) w_{ij} \quad (3.6)$$

$$\frac{\partial Q}{\partial q_i} = 0 \Rightarrow q_i = \frac{\sum_j w_{ij} q_j}{\sum_j w_{ij}} \quad (3.7)$$

Theo phương pháp biến phân thì  $f$  chứa 2 giá trị của  $f$ , có thể được coi như  $f^k$  và  $f^{k+1}$ .

Khi có được véc tơ đặc trưng  $f$  rồi, ta phân vùng cho véc tơ  $f$  thành  $C$  cụm bằng cách sử dụng một số thuật toán cơ bản như K-means hoặc dùng phương pháp ngưỡng như sau:

$$L_i = \begin{cases} 0 & \text{nếu } f_i < T_1 \\ \dots & \\ c & \text{nếu } T_c < f_i < T_{c+1} \\ \dots & \\ C & \text{nếu } f_i > T_c \end{cases}$$

Với  $T_c$  là ngưỡng thứ  $c$ .

Từ đó, ta có thuật toán **IMC** giải quyết vấn đề (3.4) như sau:

---

### Thuật toán phân cụm IMC

---

**Input:** X

**Output:** c cụm:  $T_1, T_2, \dots, T_C$

Tính  $w_{ij}$  theo công thức (3.3), khởi tạo ngẫu nhiên cho

**Lặp:**

$$\text{Tính } f^{n+1} \text{ với } f_i^{(n+1)} = \frac{\sum_j w_{ij} f_j^{(n)}}{\sum_j w_{ij}}$$

**Cho đến khi**  $|f^{(n)} - f^{(n+1)}|$  nhỏ hơn một dung sai quy định hoặc n đã đạt số vòng lặp tối đa.

**Return**  $T_1, T_2, \dots, T_C$

---

### Thuật toán tra cứu

Thuật toán 1.3 dưới đây là mô tả thuật toán tra cứu ảnh hiệu quả sử dụng phân hoạch đồ thị (An efficient image retrieval method using a graph clustering-MGC)

---

#### Thuật toán 1.3. Thuật toán tra cứu ảnh MGC

---

**Input:**

Tập các ảnh:  $S$

Ảnh truy vấn:  $Q_{initial}$

Số các ảnh được trả về tại mỗi lần lặp:  $N$

**Output:**

Danh sách kết quả tổng hợp:  $Result(Q_{merger})$

1.  $Result(Q_{initial}) \leftarrow \langle q, d, S, N \rangle;$

3. **IMC** ( $Result(Q_{initial}, N), C, X$ )

5. **Repeat**

5.1 **for**  $i=1$  to  $C$  do

$Result(Q_{merger}) \leftarrow \langle \{q^{(1)}, q^{(2)}, \dots, q^{(c)}\}, d, S, N \rangle;$

5.3  $Relevant(Q_{merger}, M) \leftarrow Feedback(Result(Q_{merger}), N');$

**until** (User dừng phản hồi);

6. **Return**  $Result(Q_{merger});$

---

## 3.2. Thực nghiệm

### 3.2.1. Môi trường thực nghiệm

Để xác định hiệu quả của các mô hình và phương pháp đề xuất, thực nghiệm được xây dựng trên nền tảng dotNET, ngôn ngữ lập trình C#, Python

và Matlab. Cấu hình máy tính sử dụng để thực nghiệm: Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, DDRam - 16GB và hệ điều hành Windows 11 Professional.

Thực nghiệm được mô tả dưới hai dạng gồm: đồ thị và bảng biểu; trong đó, hiệu suất tra cứu về độ chính xác và phạm vi được mô tả bằng đồ thị, các bảng biểu mô tả chỉ số đánh giá trung bình và so sánh giữa các phương pháp với nhau.

### CSDL ảnh thực nghiệm SIMPLIcity



*Hình III. 2. Một số ảnh trong tập SIMPLIcity*

#### 3.2.2. Thực hiện truy vấn và đánh giá

Trong phần thực nghiệm, các tham số được lựa chọn như sau:

Hiệu quả tra cứu được đánh giá trên cơ sở dữ liệu ảnh **SIMPLIcity** gồm 1000 ảnh, tất cả các ảnh trong cơ sở dữ liệu được sử dụng để thực hiện các truy vấn.

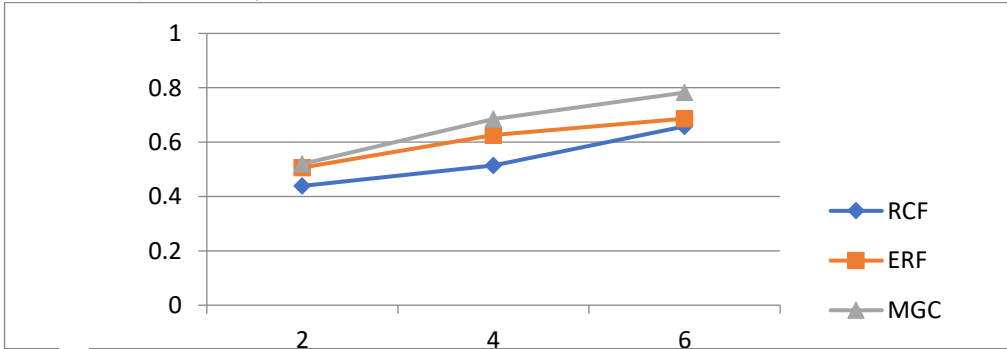
#### **So sánh độ chính xác trung bình của phương pháp đề xuất**

Năm phương pháp khác nhau gồm **CRF**, **ERF** và **MGC**.

Phương pháp	Trung bình độ chính xác		
	2	4	6
CRF	0.4388	0.5065	0.5199
ERF	0.5138	0.62658	0.6846
<b>MGC</b>	<b>0.658</b>	<b>0.68658</b>	<b>0.7825</b>

*Bảng III. 1. Bảng kết quả trung bình độ chính xác của 3 phương pháp theo số điểm truy vấn trong ba lần phản hồi.*

Trong Bảng III.1. thể hiện độ chính xác trung bình của ba phương pháp là CRF, ERF, và phương pháp đề xuất **MGC** tại các mức 2, 4 và 6 điểm truy vấn, với phương pháp đề xuất số điểm truy vấn được xác định theo số cụm. Với 2 điểm truy vấn, độ chính xác của phương pháp đề xuất cao hơn hai phương pháp CRF, ERF là 12.92%, 21.92%. Trường hợp 4 điểm truy vấn, độ chính xác của phương pháp đề xuất CRF, ERF là 12.00%, 6%. Trường hợp 8 điểm truy vấn, phương pháp đề xuất có độ chính xác cao hơn CRF, ERF lần lượt 16.47%, 26.26%.



*Hình III. 3. So sánh độ chính xác của ba phương pháp trên tập ảnh SIMPLIcity*

### Kết luận chương 3

Ngoài ra nghiên cứu sinh đã tập trung vào đề xuất phương pháp, có tên thuật toán tra cứu ảnh hiệu quả sử dụng phân hoạch đồ thị (An efficient image retrieval method using a graph clustering-MGC) mà khai thác đầy đủ thông tin độ tương tự của tập ảnh. Kết quả thực nghiệm của nghiên cứu sinh trên cơ sở dữ liệu đặc trưng ảnh đã chỉ ra rằng phương pháp được đề xuất *MGC* cung cấp một độ chính xác cao hơn hẳn so với các phương pháp khác.

### KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Luận án trình bày phương pháp ODLDA, một kỹ thuật tra cứu ảnh hiệu quả kỹ thuật cải thiện hiệu suất của hệ thống tra cứu ảnh đa điểm. ODLDA khai thác hiệu quả thông tin của người dùng thông qua tập mẫu có liên quan và không liên quan, thực hiện học phép chiếu tối ưu để tách các ảnh không liên quan và thu hẹp khoảng cách của các ảnh liên quan. Phương pháp được đề xuất tìm ma trận trọng số tối ưu của hàm khoảng cách Mahalanobis và sử dụng hàm khoảng cách cải tiến này để xếp hạng toàn bộ tập ảnh cơ sở dữ liệu và trả về tập kết quả cho người dùng. Kết quả thử nghiệm trên hai cơ sở dữ liệu đã chứng minh rằng ODLDA cung cấp độ chính xác cao hơn nhiều so với phương pháp Euclid, Euclid, RCA và OASIS cải tiến.

Ngoài ra nghiên cứu sinh đã tập trung vào đề xuất phương pháp, có tên thuật toán tra cứu ảnh hiệu quả sử dụng phân hoạch đồ thị (An efficient image retrieval method using a graph clustering-MGC) mà khai thác đầy đủ thông tin độ tương tự của tập ảnh. Kết quả thực nghiệm của nghiên cứu sinh trên cơ sở dữ liệu đặc trưng ảnh đã chỉ ra rằng phương pháp được đề xuất *MGC* cung cấp một độ chính xác cao hơn hẳn so với các phương pháp khác.

Tóm lại, luận án đã đạt một số kết quả như sau:

Thứ nhất là: Cải tiến phương pháp tra cứu ảnh thông qua tìm một phép đo khoảng cách tối ưu, mà giảm khoảng cách giữa các cặp ảnh có độ tương tự cao và tối đa hóa khoảng cách giữa các cặp ảnh có độ tương tự thấp.  
Thứ hai là: Đề xuất phương pháp tra cứu ảnh dựa trên lý thuyết cắt đồ thị, mà không phải tính ma trận Laplacian, các giá trị riêng và các véc tơ riêng.

Tuy nhiên, luận án còn một số hạn chế: phương pháp giải quyết trong luận án mới được đánh giá trên cơ sở dữ liệu vừa mà chưa xem xét trên các cơ sở dữ liệu lớn.

Từ những hạn chế trên hướng nghiên cứu tiếp theo của luận án là: tích hợp với mô hình học sâu để thích hợp với cơ sở dữ liệu lớn và tăng độ chính xác.

### **Danh mục các công trình của tác giả**

Quynh Dao Thi Thuy, Phuong Nguyen Thi Lan, Tao Ngo Quoc, Minh-Huong Ngo, ***“Improve The Efficiency Of Content-based Image Retrieval Through Incremental Clustering”***

Journal of Information Hiding and Multimedia Signal Processing, Vol. 11, No. 3, pp. 103-115, September 2020.

Jeng-Shyang Pan, Thi-Lan-Phuong Nguyen, Truong-Giang Ngo, Thi-Kien Dao, Thi-Thanh-Tan Nguyen, Trong-The Nguyen, ***“An Optimizing Cross-Entropy Thresholding for Image Segmentation based on Improved Cockroach Colony Optimization”*** JIHMSPP, Vol.11, No.4, 2020.(Scopus)

Phuong Nguyen Thi Lan, Tao Ngo Quoc, Quynh Dao Thi Thuy, Minh-Huong Ngo, ***“Improve the Effectiveness of Image Retrieval by Combining the Optimal Distance and Linear Discriminant Analysis”***

International Journal of Advanced Computer Science and Applications, <https://dx.doi.org/10.14569/IJACSA.2021.0120206>, 2021(Scopus)

The – Anh Pham<sup>1</sup>, Dinh – Nghiep Le, Thi – Lan Phuong Nguyen, ,  
**“PRODUCT SUB-VECTOR QUANTIZATION FOR FEATURE INDEXING”** Journal of Computer Science and Cybernetics, V.35, N.1  
 (2019), 69–83 DOI 10.15625/1813-9663/35/1/13442

Nguyễn Thị Lan Phương, Đỗ Văn Hải, Hoàng Văn Hùng, Trần Phạm Văn  
 Cương, **“Xây dựng cơ sở dữ liệu tổng hợp phục vụ phát triển du lịch  
 tỉnh Lào Cai bằng công nghệ GIS và viễn thám”** Tạp chí Khoa học và  
 Công nghệ Đại học Thái Nguyên. Tập 225, số 07/1, 2020

Hà Mạnh Toàn, Nguyễn Văn Năng, Trịnh Hiền Anh, Nguyễn Thị Lan  
 Phương, **“Một số kỹ thuật phân lớp người sử dụng mạng nơron tích  
 chập”**, Hội thảo quốc gia lần thứ XXI, Một số vấn đề chọn lọc của công  
 nghệ thông tin và truyền thông.

Trần Sơn Hải, Lê Quang Thái, Lê Hoàng Thái, Ngô Quốc Tạo, Nguyễn  
 Thị Lan Phương, **“Phương pháp kết hợp TLD và CMT cho theo vết đối  
 tượng chuyển động”**, Hội thảo quốc gia lần thứ XXI, Một số vấn đề chọn  
 lọc của công nghệ thông tin và truyền thông.

Nguyễn Thị Lan Phương, Đào Thị Thúy Quỳnh, Ngô Quốc Tạo, Nguyễn  
 Ngọc Quỳnh, Lê Phú Hưng, **“Nâng cao hiệu quả tra cứu ảnh dựa vào nội  
 dung sử dụng phân hoạch đồ thị “**, Hội thảo quốc gia lần thứ XXV, Một  
 số vấn đề chọn lọc của công nghệ thông tin và truyền thông. Hà Nội, ngày  
 8-9/12/2023. Nhà xuất bản Khoa học Kỹ thuật, trang 129-134 , 2022.