

# TÌM KIẾM ẢNH DỰA TRÊN ĐỒ THỊ MÔ TẢ ĐẶC TRƯNG THỊ GIÁC

Nguyễn Văn Thịnh<sup>1,\*</sup>, Đinh Thị Mận<sup>1</sup>,  
Nguyễn Thế Hữu<sup>1</sup>, Lê Trung Thu<sup>2</sup>, Văn Thế Thành<sup>1</sup>

<sup>1</sup>Trường Đại học Công nghiệp Thực phẩm TP.HCM

<sup>2</sup>Trường Đại học An Giang

\*Email: [thinhnv@hufi.edu.vn](mailto:thinhnv@hufi.edu.vn)

Ngày nhận bài: 21/01/2019; Ngày chấp nhận đăng: 06/3/2019

## TÓM TẮT

Nghiên cứu này xây dựng hệ truy vấn ảnh theo nội dung (CBIR – Content-Based Image Retrieval) và nâng cao hiệu quả tìm kiếm bằng việc đề xuất cấu trúc đồ thị mô tả đặc trưng thị giác của hình ảnh. Trước hết, nhóm tác giả đề xuất cấu trúc đồ thị lưu trữ đặc trưng thị giác của hình ảnh và độ đo tương tự giữa 2 hình ảnh dựa trên đặc trưng thị giác cấp thấp cùng với quan hệ không gian giữa các vùng trong ảnh; sau đó, trên cơ sở thuật toán K-Means, đề xuất một số cải tiến thực hiện gom cụm các đồ thị đặc trưng dựa trên độ đo tương tự. Từ đó, đề xuất thuật toán tra cứu ảnh tương tự dựa trên các cụm đồ thị. Nhằm chứng minh cơ sở lý thuyết đã đề xuất, nhóm tác giả xây dựng thực nghiệm trên tập dữ liệu ảnh ImageCLEF và đánh giá tính hiệu quả so với các phương pháp khác đã được công bố gần đây. Kết quả thực nghiệm cho thấy phương pháp được đề xuất của nghiên cứu này là hiệu quả và có thể ứng dụng trong các hệ thống dữ liệu đa phương tiện.

*Từ khóa:* Tìm kiếm ảnh theo nội dung, phân cụm, đặc trưng thị giác, độ đo tương tự.

## 1. GIỚI THIỆU

Dữ liệu số trên toàn cầu đã được gia tăng không ngừng và đạt đến một khối lượng khổng lồ. Năm 2015, tổng số hình ảnh toàn cầu đạt 3,2 nghìn tỷ; năm 2016, có 3,5 triệu ảnh được chia sẻ trong mỗi phút và có 2,5 nghìn tỷ ảnh được chia sẻ và lưu trữ trực tuyến. Trong năm 2017, thế giới đã tạo ra 1,2 nghìn tỷ ảnh và tổng số ảnh toàn cầu đến năm 2017 là 4,7 nghìn tỷ; trong đó, các hình ảnh được tạo ra từ smartphone và các thiết bị di động là 90% [1, 2]. Bên cạnh đó, ảnh số đã được ứng dụng trong nhiều hệ thống tra cứu thông tin đa phương tiện như hệ thống thông tin bệnh viện, hệ thống thông tin địa lý, hệ thống thư viện số, ứng dụng y sinh, giáo dục đào tạo, giải trí,...[3, 4]. Vì vậy, bài toán khai phá dữ liệu ảnh và tra cứu các thông tin liên quan đến hình ảnh cần được quan tâm giải quyết, tìm kiếm ảnh tương tự theo nội dung là một trong những bài toán quan trọng của các hệ thống đa phương tiện.

Nhiều hệ thống CBIR đã được phát triển như QBIC, Photobook, Visual-Seek, MARS, El Nino, CIRES, PicSOM, PicHunter, MIRROR, Virage, Netra, SIMPLITcity,... Các công trình nghiên cứu về truy vấn ảnh tương tự cũng được phát triển như truy vấn ảnh dựa trên hình dạng (shape), màu sắc (color), cấu trúc (texture), đối tượng đặc trưng (interest objects),... [5, 6]. Hầu hết các công trình tập trung vào kỹ thuật trích chọn đặc trưng, đối sánh và tìm kiếm dựa trên đặc trưng mà chưa tập trung nghiên cứu cấu trúc dữ liệu lưu trữ nhằm giảm không gian xử lý và tăng tốc độ truy vấn. Bài báo này sẽ thực hiện việc xây dựng cấu trúc dữ liệu đồ thị mô tả nội dung của hình ảnh, đồng thời xây dựng thuật toán phân hoạch cụm đồ thị và truy vấn ảnh tương tự dựa trên đồ thị. Đóng góp của bài báo gồm: (1) xây dựng cấu trúc dữ liệu đồ thị mô tả

đặc trưng thị giác của hình ảnh; (2) đề xuất độ đo đánh giá độ tương tự giữa 2 hình ảnh; (3) cải tiến thuật toán K-Means áp dụng cho việc gom cụm các đồ thị; (4) thiết kế mô hình thực nghiệm trên cơ sở cấu trúc dữ liệu, giải thuật và độ đo tương tự đã đề xuất.

Phần còn lại của bài báo gồm những nội dung như sau: Khảo sát và phân tích các công trình liên quan nhằm minh chứng tính khả thi của bài toán tìm kiếm ảnh và khắc phục các nhược điểm của các công trình đã công bố; Trình bày đồ thị mô tả đặc trưng thị giác và độ đo đánh giá mức độ tương tự giữa 2 hình ảnh nhằm thực hiện quá trình tìm kiếm ảnh tương tự; Trình bày thuật toán cải tiến phân hoạch cụm và thuật toán tra cứu ảnh để nâng cao hiệu suất tìm kiếm ảnh tương tự; Mô tả thực nghiệm và đánh giá hiệu quả của phương pháp đề xuất nhằm minh chứng tính đúng đắn của cơ sở lý thuyết; Kết luận và hướng phát triển.

## 2. CÁC CÔNG TRÌNH LIÊN QUAN

Phương pháp gom cụm đã được ứng dụng trong nhiều hệ thống tìm kiếm ảnh như áp dụng thuật toán K-means và khoảng cách Euclid để gom cụm đặc trưng màu sắc hình ảnh cho tìm kiếm ảnh [7]; tìm kiếm ảnh dựa trên thuật toán K-means và chữ ký nhị phân của hình ảnh [8]; tìm kiếm ảnh dựa trên thuật toán K-means và khoảng cách Mahalanobis giữa các vectơ màu sắc của các hình ảnh [9]; áp dụng thuật toán K-means và dải màu MPEG7 cho việc gom cụm và tìm kiếm ảnh [10]; thực hiện gom cụm kết hợp đặc trưng màu sắc, hình dạng và cấu trúc hình ảnh để tìm kiếm ảnh tương tự [11];... Theo các công trình đã công bố gần cho thấy phương pháp tìm kiếm ảnh sử dụng kỹ thuật gom cụm là khả thi và đạt kết quả tốt.

Hernández-Gracidias C.A. *et al.* (2013) thực hiện nâng cao hiệu suất tìm kiếm ảnh dựa trên quan hệ không gian (spatial relations) giữa các vùng trong ảnh. Hình ảnh được biểu diễn bằng đồ thị khái niệm dựa trên đặc trưng quan hệ không gian, pha tìm kiếm ảnh thực hiện so khớp trên toàn bộ cơ sở dữ liệu ảnh và trả về top – k ảnh liên quan nhất dựa vào độ đo [12]. Mô hình này không sử dụng đặc trưng cấp thấp biểu diễn nội dung của hình ảnh để tăng độ chính xác, bên cạnh đó thời gian tra cứu ảnh sẽ chậm do phải so khớp tuyến tính trong toàn bộ cơ sở dữ liệu ảnh.

Hakan Cevikalp *et al.* (2017) thực hiện tra cứu ảnh dựa trên cấu trúc đồ thị và cây phân cấp nhị phân. Quá trình huấn luyện được thực hiện theo phương pháp SVM (Support Vector Machines) dựa trên các đặc trưng hình ảnh cấp thấp [14]. Phương pháp đã thực nghiệm trên bộ ảnh ImageCLEF và so sánh tính hiệu quả với các phương pháp khác.

Thanh The Van *et al.* (2018) đề xuất hệ tìm kiếm ảnh theo nội dung dựa trên đồ thị cụm chữ ký nhị phân. Tác giả thực hiện gom cụm chữ ký nhị phân mô tả đặc trưng thị giác cấp thấp của ảnh và tạo đồ thị cụm lưu trữ các chữ ký nhị phân này nhằm nâng cao hiệu quả tìm kiếm ảnh. Phương pháp đã thực nghiệm trên nhiều bộ dữ liệu và so sánh tính hiệu quả với nhiều phương pháp khác nhau [15].

Trên cơ sở kế thừa các công trình đã có và khắc phục những hạn chế của các phương pháp liên quan đã công bố, đồng thời tạo ra một hệ truy vấn ảnh theo nội dung nhằm nâng cao hiệu quả tìm kiếm, nghiên cứu này đề xuất một mô hình truy vấn ảnh bằng cách tạo đồ thị kết hợp đặc trưng thị giác cấp thấp và quan hệ không gian của hình ảnh, sau đó gom cụm các đồ thị đặc trưng thị giác nhằm nâng cao hiệu suất cho việc tìm kiếm ảnh tương tự.

## 3. ĐỒ THỊ MÔ TẢ VECTƠ ĐẶC TRƯNG VÀ ĐỘ ĐO TƯƠNG TỰ

### 3.1. Đồ thị mô tả vectơ đặc trưng thị giác

Mỗi hình ảnh trong tập dữ liệu được phân đoạn thành nhiều vùng khác nhau theo phương pháp của Hugo Jair Escalante *et al.* [16]. Hình 1 mô tả một ảnh gốc và 5 ảnh của các

vùng thuộc về các lớp ảnh: cloud (1), hill (2), ruin-archeological (208), road (203), group-of-persons (120) của ảnh 9502.jpg. Mỗi vùng được trích xuất một vector đặc trưng bao gồm đặc trưng vùng: diện tích, chiều rộng và chiều cao; đặc trưng về vị trí: giá trị trung bình và độ lệch chuẩn theo trục x và trục y; đặc trưng về hình dạng (shape): boundary/area, convexity; đặc trưng màu sắc trong không gian RGB và CIE-Lab: trung bình, độ lệch chuẩn và độ nghiêng.



Hình 1. Ảnh gốc và các ảnh phân đoạn (9502.jpg)

Để lưu trữ các hình ảnh theo mô tả như trên, nhóm tác giả đề xuất đồ thị mô tả vector đặc trưng thị giác như sau:

Đồ thị mô tả vector đặc trưng thị giác (VFG) là một đồ thị vô hướng, mô tả một hình ảnh trong tập dữ liệu ảnh. Mỗi đỉnh trong đồ thị chứa cặp giá trị  $(id_I, f_I)$ , trong đó  $id_I, f_I$  lần lượt là định danh và đặc trưng thị giác của vùng  $I$  trong một ảnh.

$VFG = (V, E)$ , trong đó tập các đỉnh  $V$  và tập các cạnh  $E$  được định nghĩa:

$$V = \{v_i | v_i = (id(I_i), f(I_i), I_i \in I, i = 1..m)\}$$

$$E = \{(v_i, v_j) | \phi(v_i, v_j) \in \{1,2\}; i, j = 1..m\}$$

Với  $I_i, I_j$  là vùng ảnh thứ  $i, j$  trong  $m$  vùng ảnh của  $I$ ;  $\phi(v_i, v_j) = 1$  nếu hai đỉnh  $v_i$  và  $v_j$  kề nhau, ngược lại thì  $\phi(v_i, v_j) = 2$ . Hình 2 minh họa đồ thị mô tả vector thị giác của ảnh 9502.jpg trong tập dữ liệu ảnh.

Cấu trúc dữ liệu đồ thị mô tả vector đặc trưng thị giác được định nghĩa:

```
struct vertex{
    int region_id;//id của vùng
    vector<double> region_features;//đặc trưng của vùng
    int region_label;//phân lớp của vùng
};
struct graph{
    int image_id; //id của hình ảnh
    vector<vertex> lstRegions;//tập đỉnh
    matrix<int> adj_regions;//tập cạnh
};
```

Image\_id: 9502

Tập đỉnh

Region id	Features							Label
1	0.09887731	0.2277778	0.3852076	0.08225318	1.093066	0.1992732	...	60
2	0.4842998	0.8194444	0.5439445	0.348718	1.03805	0.5856359	...	131
3	0.1932755	0.4333333	0.4137344	0.6962593	0.9543005	0.4024132	...	208
4	0.09413773	0.1666667	0.5687629	0.9433828	1.10365	0.1407718	...	203
5	0.1188657	0.3972222	0.5032548	0.7950058	0.3397966	0.3443138	...	120

Tập cạnh

	1	2	3	4	5
1	2	1	2	2	2
2	1	2	1	2	2
3	2	1	2	1	1
4	2	2	1	2	1
5	2	2	1	1	2

Hình 2. Đồ thị vec-tơ đặc trưng thị giác của ảnh 9502.jpg trong tập dữ liệu ảnh

### 3.2. Độ đo tương tự giữa hai đồ thị

Độ đo tương tự giữa 2 hình ảnh là thành phần không thể thiếu khi thực hiện quá trình gom cụm và truy vấn các hình ảnh tương tự. Mỗi hình ảnh được mô tả dưới dạng một đồ thị, do đó vấn đề đặt ra là xây dựng độ đo tương tự giữa 2 đồ thị để từ đó đánh giá độ tương tự giữa các hình ảnh. Dựa trên đồ thị đặc trưng của hình ảnh, bài báo định nghĩa độ đo tương tự giữa các hình ảnh như sau:

**Định nghĩa 3.2.1:** Độ tương tự đặc tính giữa hai hình ảnh

Cho 2 tập vectơ đặc tính của ảnh  $I$  và  $J$  lần lượt là  $V_I = \{v_i^i \mid i = 1, \dots, n\}$  và  $V_J = \{v_j^j \mid j = 1, \dots, m\}$ . Độ đo đặc tính giữa một vectơ đặc tính  $v_i^i$  của ảnh  $I$  với ảnh  $J$  được đánh giá theo công thức như sau:

$$f(v_i^i, J) = \min\{|v_i^i - v_j^j|, v_j^j \in V_J, j = 1 \dots m\} \quad (1)$$

Từ công thức (1), độ đo tương tự đặc tính giữa 2 hình ảnh được đánh giá theo công thức sau:

$$F(I, J) = \sum_{i=1}^n f(v_i^i, J) + \sum_{j=1}^m f(v_j^j, I) \quad (2)$$

**Định lý 3.2.1:** Hàm đánh giá độ đo  $F(I, J)$  là một khoảng cách trong không gian định chuẩn vì có các tính chất sau:

- (1) Không âm:  $F(I, J) \geq 0$  và  $F(I, J) = 0 \Leftrightarrow V_I = V_J$
- (2) Đối xứng:  $F(I, J) = F(J, I)$
- (3) Bất đẳng thức tam giác:  $F(I, J) + F(J, K) \geq F(I, K)$

**Chứng minh:**

#### (1) Tính không âm

Gọi  $V_I$  và  $V_J$  là hai vectơ đặc tính của hai hình ảnh  $I$  và  $J$  bất kỳ.

Vì  $|v_i^i - v_j^j| \geq 0$  nên  $f(v_i^i, J) \geq 0$  và  $f(v_j^j, I) \geq 0$ . Khi đó  $\sum_{i=1}^n f(v_i^i, J) \geq 0$  và  $\sum_{j=1}^m f(v_j^j, I) \geq 0$ .

Suy ra  $F(I, J) \geq 0$ .

Vì vậy, hàm đánh giá độ tương tự  $F(I, J)$  có tính không âm.

#### (2) Tính đối xứng

Vì phép toán  $+$  có tính chất giao hoán, nên:

$$F(I, J) = \sum_{i=1}^n f(v_i^i, J) + \sum_{j=1}^m f(v_j^j, I) = \sum_{j=1}^m f(v_j^j, I) + \sum_{i=1}^n f(v_i^i, J) = F(J, I)$$

Do đó,  $F(I, J) = F(J, I)$

Vì vậy, hàm đánh giá độ tương tự  $F(I, J)$  có tính chất đối xứng.

#### (3) Bất đẳng thức tam giác

Gọi  $V_I, V_J$  và  $V_K$  là lần lượt là 3 vectơ đặc tính của 3 hình ảnh  $I, J, K$  bất kỳ. Khi đó:

$$F(I, J) + F(J, K) = \sum_{i=1}^n f(v_i^i, J) + \sum_{j=1}^m f(v_j^j, I) + \sum_{j=1}^m f(v_j^j, K) + \sum_{k=1}^o f(v_k^k, J)$$

Dễ thấy:

$$\sum_{i=1}^n f(v_i^i, J) + \sum_{j=1}^m f(v_j^j, I) + \sum_{j=1}^m f(v_j^j, K) + \sum_{k=1}^o f(v_k^o, J) \geq \sum_{i=1}^n f(v_i^j, K) + \sum_{k=1}^o f(v_k^o, I)$$

Do đó:  $F(I, J) + F(J, K) \geq F(I, K)$

Vì vậy, hàm đánh giá độ đo tương tự thỏa bất đẳng thức tam giác.

**Định nghĩa 3.2.2:** Độ đo không gian giữa hai đồ thị

Cho 2 tập bậc các đỉnh kề của ảnh  $I$  và  $J$  lần lượt là  $D_I = \{d_i \mid i = 1, \dots, n\}$  và  $D_J = \{d_j \mid j = 1, \dots, m\}$ . Độ đo không gian giữa một vùng của ảnh  $I$  với ảnh  $J$  được đánh giá theo công thức như sau:

$$s(d_i, J) = \min\{|d_i - d_j|, d_j \in D_J, j = 1, \dots, m\} \quad (3)$$

Từ công thức (3), độ đo tương tự không gian giữa 2 hình ảnh được đánh giá theo công thức sau:

$$S(I, J) = \sum_{i=1}^n s(d_i, J) + \sum_{j=1}^m s(d_j, I) \quad (4)$$

**Định lý 3.2.2:** Hàm đánh giá độ đo  $S(I, J)$  là một khoảng cách trong không gian định chuẩn vì có các tính chất sau:

- (1) Không âm:  $S(I, J) \geq 0$  và  $S(I, J) = 0 \Leftrightarrow D_I = D_J$
- (2) Đối xứng:  $S(I, J) = S(J, I)$
- (3) Bất đẳng thức tam giác:  $S(I, J) + S(J, K) \geq S(I, K)$

**Chứng minh:** Dễ thấy  $S(I, J)$  là một khoảng cách trong không gian định chuẩn.

Dựa trên công thức (2) và (4). Độ đo tương tự giữa 2 ảnh được tính theo công thức (5) như sau:

$$sim(I, J) = \alpha F(I, J) + \beta S(I, J); \alpha + \beta = 1; \alpha, \beta \in (0, 1) \quad (5)$$

## 4. THUẬT TOÁN PHÂN HOẠCH CỤM VÀ TÌM KIẾM ẢNH

### 4.1. Thuật toán phân hoạch cụm đồ thị theo độ đo tương tự

Trong thuật toán K-means, 3 tham số đầu vào cần phải có gồm số cụm, tâm cụm và độ đo. Bên cạnh đó, khi một phần tử mới được thêm vào cụm thì tâm của cụm chứa phần tử đó phải được xác định lại. Việc này không phù hợp đối với các bài toán dữ liệu kích thước lớn vì sẽ dẫn tới 2 phần tử trong cùng một cụm có khoảng xa nhau, đồng thời quá trình cập nhật lại tâm cụm mất nhiều thời gian ảnh hưởng đến hiệu suất của hệ thống. Do đó, trong bài báo này, nhóm tác giả cải tiến thuật toán K-means để áp dụng cho việc gom cụm đồ thị mô tả đặc trưng thị giác của hình ảnh như sau: (1) số cụm tăng trưởng tùy theo số đồ thị và độ tương tự giữa các đồ thị; (2) xác định ngưỡng độ đo giữa 2 đồ thị trong quá trình gom cụm và cố định tâm cụm theo ngưỡng.

#### Thuật toán GraphClustering

**Đầu vào:** Tập đồ thị đặc trưng  $G$  và ngưỡng  $\theta$ .

**Đầu ra:** Tập cụm  $C$ .

**Function** GraphClustering( $G, \theta$ )

    Khởi tạo  $C = \phi$ ;

**for**  $g \in G$  **do**

```

if  $C = \phi$  then
    Khởi tạo cụm  $C_0$  với tâm  $g$ ;
else
    Tìm cụm  $C_k \in C: sim(g, centroid_k) = \min\{sim(g, centroid_i), i = 1, \dots, m\}$ 
    //centroidk là tâm của cụm k
    //sim(g, gi): độ tương tự của đồ thị gi và g;
    if  $sim(g, centroid_k) < \theta$  then
         $C_k = C_k \cup g$ ;
    else
        Tạo cụm  $C_i$  mới với tâm là  $g$ ;
         $C = C \cup C_i$ ;
    end if
end if
end for
return  $C$ ;

```

**End.**

**Mệnh đề 1.** Độ phức tạp của **Thuật toán GraphClustering** là  $O(N^2)$ , với  $N$  là số đồ thị trong tập đồ thị  $G$ .

**Chứng minh.** Cho  $N$  là số phần tử của tập  $G$ . Dòng lệnh “for  $g \in G$ ” do thực hiện  $N$  lần, ứng với mỗi lần thực hiện, dòng lệnh “Tìm cụm  $C_k \in C: sim(g, centroid_k) = \min\{sim(g, centroid_i), i = 1, \dots, m\}$ ” thực hiện  $m$  phép toán để tìm được cụm có tâm gần nhất, với  $m$  là số cụm ở thời điểm hiện hành. Vì vậy, số phép toán của **Thuật toán GraphClustering** là  $\sum_{m=1}^N \sum_{i=1}^m H = N^2 \times H$ , với  $H$  là hằng số. Do đó, độ phức tạp **Thuật toán GraphClustering** là  $O(N^2)$ .

#### 4.2. Thuật toán tìm kiếm ảnh tương tự

Trên cơ sở tập cụm kết quả ở pha tiền xử lý, nhóm tác giả thực hiện tra cứu ảnh bằng cách chọn cụm có tâm gần với ảnh truy vấn nhất. Nếu khoảng cách từ ảnh truy vấn tới tâm cụm nhỏ hơn ngưỡng  $\theta$  thì lấy ra tập các ảnh trong cụm và sắp xếp theo độ đo tương tự. Thuật toán tra cứu ảnh được thực hiện như sau:

**Thuật toán ImageRetrieval**

**Đầu vào:** Đồ thị đặc trưng của ảnh truy vấn  $g$ , tập cụm (cluster)  $C$  và ngưỡng  $\theta$ .

**Đầu ra:** Tập ảnh tương tự  $I$ .

**Function** ImageRerieval( $g, C, \theta$ )

Khởi tạo  $I = \phi$ ;

Tìm cụm  $C_k \in C: sim(g, cen_k) = \min\{sim(g, cen_i), i = 1, \dots, m\}$ ;

**if**  $sim(g, cen_k) < \theta$  **then**

$I = I \cup g_i, g_i \in C_k, i = 1, \dots, |C_k|$ ;

**end if**

Sắp xếp tập  $I$  giảm dần theo  $sim(g, cen_i)$ ;

**return**  $I$ ;

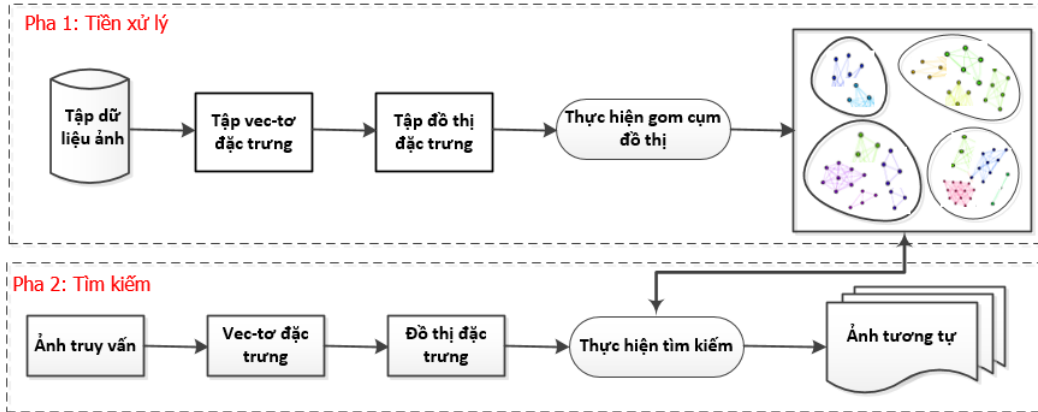
**End.**

**Mệnh đề 2.** Độ phức tạp của **Thuật toán ImageRetrieval** là  $O(n)$ , với  $n$  là số lượng cụm trong tập  $C$ .

**Chứng minh.** Gọi tập cụm  $C$  có số lượng cụm là  $m$ . Với mỗi ảnh (đồ thị) cần truy vấn, thuật toán tìm cụm có tâm gần với đồ thị truy vấn nhất trong tập  $C$  dựa trên độ đo tương tự. Thuật toán thực hiện duyệt tối đa là  $n$  lần để tìm cụm phù hợp và trích xuất tập các đồ thị tương ứng. Do đó, độ phức tạp của **Thuật toán ImageRetrieval** là  $O(n)$ .

## 5. XÂY DỰNG ỨNG DỤNG THỰC NGHIỆM

### 5.1. Mô hình ứng dụng thực nghiệm



Hình 3. Mô hình hệ thống tìm kiếm ảnh dựa trên đồ thị mô tả vec-tơ đặc trưng thị giác

Mô hình tổng quát của hệ thống tìm kiếm ảnh theo nội dung dựa trên đồ thị mô tả vectơ đặc trưng thị giác được mô tả tại Hình 3. Mô hình tìm kiếm ảnh gồm 2 pha: tiền xử lý và tìm kiếm ảnh tương tự.

#### Tiền xử lý:

**Bước 1:** tạo tập đồ thị từ tập vector đặc trưng thị giác của từng phân vùng của mỗi hình ảnh, mỗi đồ thị biểu diễn một hình ảnh trong tập dữ liệu ảnh;

**Bước 2:** đánh giá độ tương tự giữa 2 đồ thị giữa trên độ đo tương tự đề xuất;

**Bước 3:** gom cụm các đồ thị theo độ đo tương tự.

#### Tìm kiếm ảnh tương tự:

**Bước 1:** từ ảnh truy vấn, tạo đồ thị mô tả đặc trưng thị giác của ảnh này;

**Bước 2:** thực hiện truy vấn trên các cụm để cho kết quả là cụm tương tự với ảnh truy vấn nhất;

**Bước 3:** kết xuất các ảnh kết quả sắp xếp theo độ đo tương tự với ảnh truy vấn.

### 5.2. Kết quả thực nghiệm

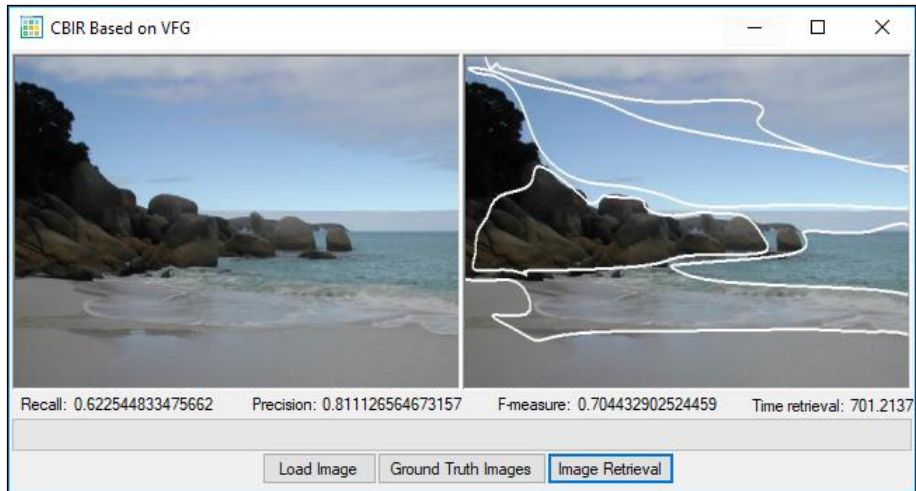
Thực nghiệm trên máy PC CPU Intel Core i5-7200U CPU @ 2.50GHz, 8.0GB RAM, hệ điều hành Windows 10 Pro 64 bit. Kết quả thực nghiệm được đánh giá trên bộ dữ liệu imageCLEF chứa 20.000 ảnh, được chia thành 276 lớp và lưu trữ trong 41 thư mục (từ thư mục 0 đến thư mục 40); bộ dữ liệu của kích thước 1,64 GB. Để đánh giá hiệu quả của phương pháp tìm kiếm ảnh, phần thực nghiệm được đánh giá các giá trị gồm: độ chính xác (precision), độ phủ (recall) và độ đo dung hòa F-measure. Công thức tính các giá trị này như sau [4]:

$$precision = \frac{|relevant\ images \cap retrieved\ images|}{|retrieved\ images|} \quad (6)$$

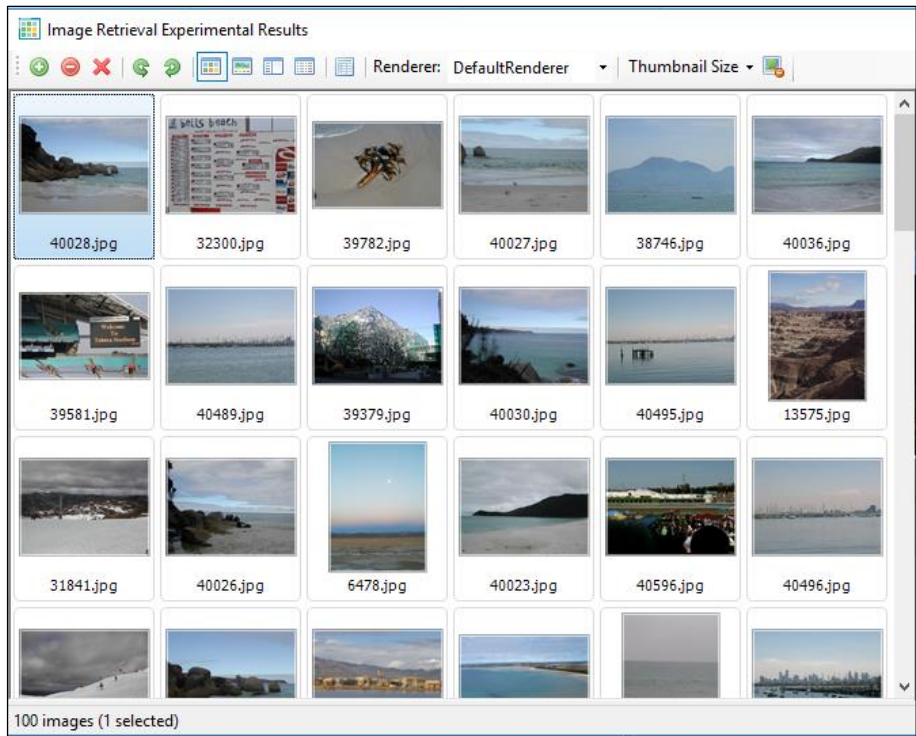
$$recall = \frac{|relevant\ images \cap retrieved\ images|}{|retrieved\ images|} \quad (7)$$

$$F\text{-measure} = 2 \times \frac{(precision \times recall)}{(precision + recall)} \quad (8)$$

Trong đó: *relevant images* là tập ảnh tương tự với ảnh truy vấn và có trong tập dữ liệu ảnh, *retrieved images* là tập ảnh đã tìm kiếm được. Các giá trị độ chính xác, độ phủ và độ đo dung hòa được tính theo tỷ lệ % và được quy đổi thành giá trị trên đoạn [0, 1].



Hình 4. Ứng dụng thực nghiệm của phương pháp đề xuất



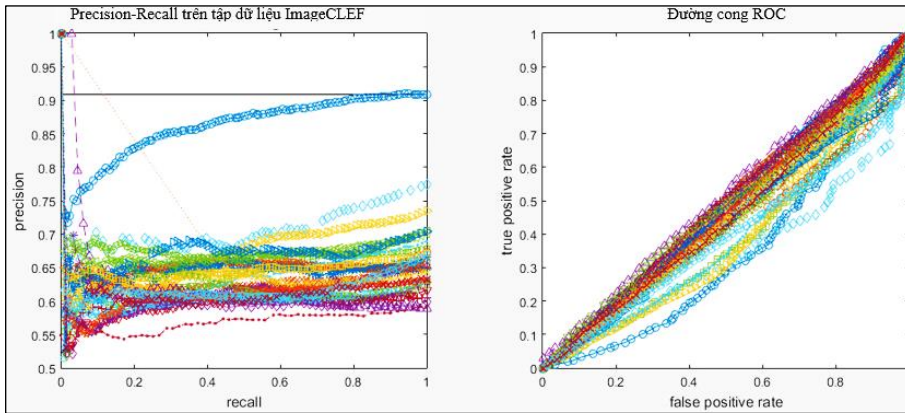
Hình 5. Kết quả tìm kiếm ảnh của phương pháp đề xuất



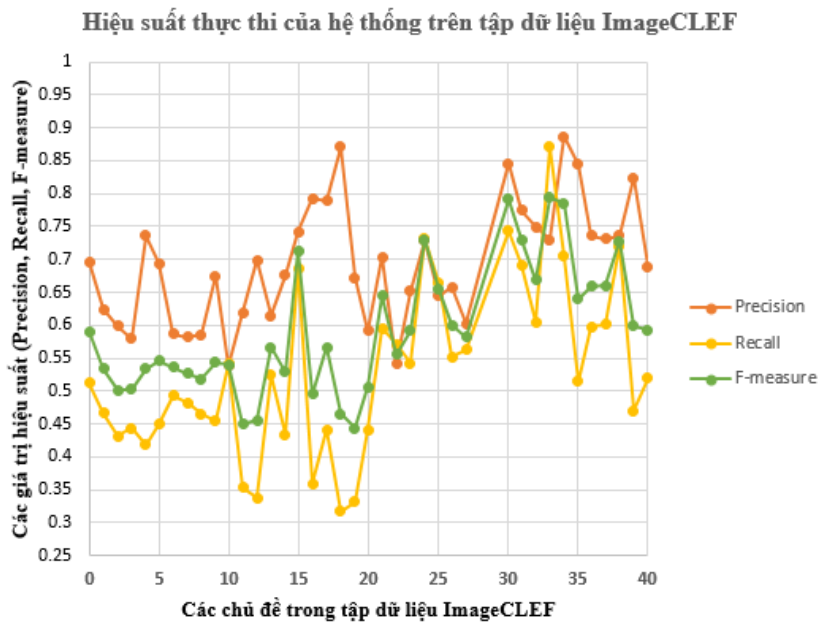
Bảng 1. Hiệu suất tìm kiếm ảnh của phương pháp được đề xuất trên bộ dữ liệu ImageCLEF

Tập ảnh	Số ảnh	Số cụm	Độ chính xác trung bình	Độ phủ trung bình	Độ đo dung hòa trung bình
00-10	6450	6	0,624761	0,460895	0,530461
11-20	4857	6	0,708162	0,442224	0,554454
21-30	3615	4	0,642329	0,622906	0,632468
31-40	5078	4	0,776099	0,630069	0,695501

Kết quả thực nghiệm của phương pháp được mô tả tại Hình 5, Hình 6; hiệu suất của phương pháp đề xuất được trình bày trong Bảng 1; giá trị độ chính xác truy vấn trung bình (MAP) của phương pháp đề xuất được so sánh với các phương pháp khác trên cùng bộ dữ liệu ImageCLEF được mô tả trong Bảng 2. Hình 6 mô tả giá trị precision, recall, F-measure trung bình của 39 bộ ảnh trong tập dữ liệu ImageCLEF. Kết quả tại Bảng 2 cho thấy phương pháp được đề xuất tương đối chính xác so với các hệ truy vấn ảnh theo nội dung khác.



Hình 6. Độ chính xác - Độ phủ và đường cong ROC của ứng dụng trên tập dữ liệu ImageCLEF



Hình 7. Trung bình các giá trị hiệu suất trên tập dữ liệu ImageCLEF

Bảng 2. So sánh độ chính xác giữa các phương pháp trên bộ dữ liệu ImageCLEF

Phương pháp	Độ chính xác truy vấn trung bình (MAP)
Hakan Cevikalp, 2017 [14]	0,4678
Thanh The Van, 2018 [15]	0,7945
C.A. Hernández-Gracidas, 2013 [12]	0,5826
Ji Wan, 2015 [13]	0,6632
Đề xuất của nhóm tác giả	0,6878

## 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã xây dựng được phương pháp tìm kiếm ảnh dựa trên cấu trúc đồ thị mô tả đặc trưng thị giác của hình ảnh, đề xuất độ đo tương tự giữa 2 hình ảnh và cải tiến thuật toán K-means thực hiện gom cụm đồ thị đặc trưng nhằm tăng hiệu suất tìm kiếm ảnh tương tự. Trên cơ sở lý thuyết đã được đề xuất, nhóm tác giả xây dựng hệ truy vấn ảnh theo nội dung. Kết quả thực nghiệm trên bộ dữ liệu ảnh ImageCLEF được đánh giá và so sánh với các công trình khác trên cùng một tập dữ liệu ảnh đã cho thấy phương pháp đề xuất là hiệu quả. Vì vậy, phương pháp đề xuất trong nghiên cứu này có thể được sử dụng cho các hệ thống tra cứu dữ liệu đa phương tiện. Việc cố định tâm cụm đã giảm được đáng kể thời gian của quá trình gom cụm. Tuy nhiên, việc này dẫn đến có thể xuất hiện các cụm có ít phần tử hoặc có quá nhiều phần tử, ảnh hưởng đến độ chính xác của hệ thống. Hướng phát triển tiếp theo của nghiên cứu là xây dựng thuật toán tách/gộp cụm nhằm đảm bảo các phần tử trong cùng một cụm phải tương tự nhau, đồng thời bổ sung thêm thông tin quan hệ không gian giữa các vùng trong ảnh để tăng độ chính xác.

**Lời cảm ơn:** Nghiên cứu này được Trường Đại học Công nghiệp Thực phẩm TP.HCM tài trợ và được nhóm nghiên cứu SBIR-HCM, Trường Đại học Sư phạm TP.HCM hỗ trợ về chuyên môn và cơ sở vật chất.

## TÀI LIỆU THAM KHẢO

1. John Gantz, David Reinsel - The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east, IDC iView, EMC Corporation, Tech. Rep. (2013).
2. Deloitte - Photo sharing: trillions and rising, Deloitte Touche Tohmatsu Limited, Deloitte Global (2016).
3. Muneesawang P., Zhang N., Guan L. - Multimedia Database Retrieval: Technology and Applications, Graduate Texts in Mathematics, Springer, New York Dordrecht London (2014).
4. Xie X., Cai X., Zhou J., Cao N., Wu Y. - A semantic-based method for visualizing large image collections, IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Society (2018). <https://doi.org/10.1109/TVCG.2018.2835485>.
5. Kim S., Park S., Kim M. - Central object extraction for object-based image retrieval. In: Bakker E.M., Lew M.S., Huang T.S., Sebe N., Zhou X.S. (eds) Image and Video Retrieval, CIVR 2003, Lecture Notes in Computer Science **2728**, Springer (2003) 39-49.
6. Y.-K. N. Hun-Woo Yoo, She-Hwan Jung, Dong-Sik Jang – Extraction of major object features using VQ clustering for content-based image retrieval, Pattern Recognition **35** (2002) 1115-1126.
7. Chuen-Horng Lin, Chun-Chieh Chen, Hsin-Lun Lee, Jan-Ray Liao – Fast K-means algorithm based on a level histogram for image retrieval, Expert System with Application **41** (7) (2014) 3276-3283.

8. Thanh The Van, Nguyen Van Thinh, Thanh Manh Le - The method proposal of image retrieval based on K-Means algorithm, *Advances in Intelligent Systems and Computing* **746** (2) (2018) 481-490.
9. M. Banerjee, S. Bandyopadhyay, and S. K. Pal - A clustering approach to image retrieval using range based query mahalanobis distance, *Intelligent Systems Reference Library* **43** (2013) 79-91.
10. H. R. Saboorian, M.M., Jamzad, M. Rabiee – User adaptive clustering for large image databases, 20th IEEE International Conference on Pattern Recognition, no. 11578177, (2010), pp. 4271–4274.
11. N. Zakariya, S.M., Ali, R., Ahmad - Combining visual features of an image at different precision value of unsupervised content based image retrieval, 2010 IEEE International Conference on Computational Intelligence and Computing Research, (2010) 1-4.
12. Arturo C., Hernández-Gracidas, Sucar L.E., Montes M. - Improving image retrieval by using spatial relations, *Multimedia Tools and Applications* **62** (2) (2013) 479-505.
13. X. G. i Wan, Pengcheng Wu, Steven C. H. Hoi, Peilin Zhao and J. L. Dayong Wang, Yongdong Zhang - Online learning to rank for content-based image retrieval, Twenty-Fourth International Joint Conference on Artificial Intelligence (2015).
14. Hakan Cevikalp, Merve Elmas, Savas Ozkan - Large-scale image retrieval using transductive support vector machines, *Computer Vision and Image Understanding* **173** (2018) 2-12.
15. Thanh The Van, Thanh Manh Le - Content-based image retrieval based on binary signatures cluster graph, *Expert Systems: The Journal of Knowledge Engineering* **35** (1) (2018) 1-22.
16. Hugo Jair Escalante, Carlos A. Hernández, Jesus A. Gonzalez, A. López- López, Manuel Montes, Eduardo F. Morales, L. Enrique Sucar, Luis Villasenor, Michael Grubinger - The segmented and annotated IAPR TC-12 benchmark, *Computer Vision and Image Understanding* **114** (4) (2010) 419-428.

## **ABSTRACT**

### **IMAGE RETRIEVAL BASED ON VISUAL FEATURE GRAPH**

Nguyen Van Thinh<sup>1\*</sup>, Dinh Thi Man<sup>1</sup>,  
Nguyen The Huu<sup>1</sup>, Le Trung Thu<sup>2</sup>, Van The Thanh<sup>1</sup>

<sup>1</sup>*Ho Chi Minh City University of Food Industry*

<sup>2</sup>*An Giang University*

\*Email: *thinhnv@hufi.edu.vn*

In this paper, the content-based image retrieval is built and the retrieval efficiency is improved by proposing the visual feature graph of the image. First of all, based on low-level features and spatial relations among regions in the image, a visual feature graph of images and similarity measure between two images is proposed; Then, on the base of K-Means algorithm, several improvements to cluster the visual feature graphs are proposed based on the similarity measure. Since then, an image retrieval algorithm is done on graph clusters. To illustrate the proposed theory, we build experiments on ImageCLEF dataset and assess the effectiveness of our method and compared to the ones of work which recently published on the same dataset. Experimental results show that our proposed method is effective and it can be applied in many multimedia data systems.

*Keywords:* CBIR, clustering, visual feature, similarity measure.