

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

Nguyễn Văn Thịnh\*, Đinh Thị Mận  
Nguyễn Thế Hữu, Nguyễn Thị Định, Văn Thế Thành

Trường Đại học Công nghiệp Thực phẩm TP.HCM

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

Ngày nhận bài: 09/7/2019; Ngày chấp nhận đăng: 05/9/2019

## TÓM TẮT

Bài báo xây dựng hệ truy vấn ảnh và nâng cao hiệu quả tìm kiếm bằng việc đề xuất cấu trúc đồ thị mô tả đặc trưng không gian của hình ảnh. Trước hết, nhóm tác giả đề xuất cấu trúc đồ thị lưu trữ đặc trưng không gian của hình ảnh và độ đo tương tự giữa hai hình ảnh dựa trên lớp ngữ nghĩa và quan hệ không gian giữa các vùng trong ảnh. Sau đó, trên cơ sở thuật toán K-Means, một số cải tiến thực hiện gom cụm các đồ thị đặc trưng dựa trên độ đo tương tự được đề xuất. Từ đó, đề xuất thuật toán tra cứu ảnh dựa trên các cụm đồ thị. Nhằm chứng minh cơ sở lý thuyết đã đề xuất, thực nghiệm được xây dựng và đánh giá trên tập dữ liệu ảnh ImageCLEF; kết quả thực nghiệm được đánh giá so với các phương pháp khác đã được công bố gần đây trên cùng tập dữ liệu. Từ kết quả thực nghiệm cho thấy, phương pháp đề xuất của nhóm tác giả là hiệu quả và có thể ứng dụng trong nhiều hệ thống dữ liệu đa phương tiện.

*Từ khóa:* Tìm kiếm ảnh, gom cụm, đặc trưng không gian, độ đo tương tự.

## 1. GIỚI THIỆU

Có nhiều phương pháp tìm kiếm các hình ảnh như tìm kiếm hình ảnh dựa trên từ khóa TBIR (Text-based image retrieval) được giới thiệu từ những năm 1970, tìm kiếm hình ảnh dựa trên nội dung CBIR (Content-based image retrieval) được giới thiệu từ những năm 1980, tìm kiếm ảnh dựa trên ngữ nghĩa SBIR (Semantic-based image retrieval) là bài toán được quan tâm hiện nay [1]. Nhiều công trình nghiên cứu về tìm kiếm ảnh tương tự đã được phát triển dựa trên đặc trưng cấp thấp như truy vấn ảnh dựa trên hình dạng (*shape*), màu sắc (*color*), cấu trúc (*texture*), đối tượng đặc trưng (*interest objects*)... [2]. Để giải quyết bài toán tra cứu dữ liệu ảnh, đã có nhiều hệ thống tìm kiếm ảnh được phát triển như QBIC, Photobook, Visual-Seek, MARS, El Nino, CIRES, PicSOM, PicHunter, MIRROR, Virage, Netra, SIMPLITcity... [3].

Mặt khác, 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 tỉ; năm 2016, có 3,5 triệu ảnh được chia sẻ trong mỗi phút và có 2,5 nghìn tỉ ảnh được chia sẻ và lưu trữ trực tuyến. Năm 2017, thế giới đã tạo ra 1,2 nghìn tỉ ảnh và tổng số ảnh toàn cầu đến năm 2017 là 4,7 nghìn tỉ. Năm 2019, mỗi ngày có 8,95 triệu ảnh, video được chia sẻ trên Instagram và trên toàn cầu đã có hơn 40 nghìn tỷ ảnh [4]. Ảnh số đã được sử 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, thư viện số, ứng dụng y sinh, giáo dục đào tạo, giải trí... [5]. 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.

Trong bài báo này, nhóm tác giả xây dựng hệ truy vấn ảnh dựa trên đồ thị đặc trưng không gian nhằm tăng tốc độ tìm kiếm và đảm bảo được độ chính xác. Đóng góp của bài báo

gồm: (1) xây dựng đồ thị đặc trưng không gian của hình ảnh; (2) đề xuất độ đo tương tự giữa hai đồ thị nhằm đánh giá độ tương tự giữa các hình ảnh; (3) cải tiến thuật toán K-Means áp dụng cho việc gom cụm các đồ thị; (4) đề xuất mô hình truy vấn ảnh tương tự dựa trên các cơ sở lý thuyết và thuật toán đã xây dựng; (5) xây dựng ứng dụng thực nghiệm về mô hình truy vấn ảnh trên bộ dữ liệu ImageCLEF.

Phần còn lại của bài báo gồm: Phần 2 đề cập đến các công trình nghiên cứu liên quan nhằm phân tích tính khả thi của phương pháp đề xuất; Phần 3 xây dựng cấu trúc dữ liệu đồ thị đặc trưng không gian nhằm lưu trữ dữ liệu mô tả của hình ảnh, độ đo tương tự giữa hai đồ thị đặc trưng không gian cũng được trình bày trong phần này; Phần 4 trình bày thuật toán cải tiến phân hoạch cụm, thuật toán tra cứu ảnh nhằm nâng cao hiệu suất tìm kiếm ảnh tương tự; mô hình tìm kiếm ảnh và thực nghiệm được trình bày trong Phần 5; kết luận và hướng phát triển được trình bày trong Phần 6.

## 2. CÁC CÔNG TRÌNH LIÊN QUAN

Nhiều công trình liên quan đến tìm kiếm ảnh đã được tiếp cận và công bố trong những năm gần đây. Phương pháp gom cụm đã được ứng dụng trong nhiều hệ thống tìm kiếm ảnh như áp dụng thuật toán K-means để gom cụm đặc trưng màu sắc hình ảnh [6]; tìm kiếm ảnh dựa trên thuật toán K-means và chữ ký nhị phân của hình ảnh [7]; tìm kiếm ảnh dựa trên thuật toán K-means và khoảng cách Mahalanobis giữa các véc-tơ màu sắc của các hình ảnh [8]; áp dụng thuật toán K-means và dải màu MPEG7 cho việc gom cụm và tìm kiếm ảnh [9]; thực hiện gom cụm kết hợp đặc trưng màu sắc, hình dạng và cấu trúc hình ảnh để tìm kiếm ảnh tương tự [10]. Ngoài ra, nhiều công trình nghiên cứu trích xuất ngữ nghĩa hình ảnh và tìm kiếm ảnh tương tự cũng đã được giới thiệu trong những năm gần đây như: tìm kiếm ảnh dựa trên túi từ thị giác [11], tra cứu ảnh dựa trên việc phân tích ngôn ngữ tự nhiên để tạo ra câu truy vấn SPARQL [12], xây dựng công cụ I2T (Image to Text) nhằm tạo ra các RDF mô tả ngữ nghĩa hình ảnh [13] ... Theo kết quả các công trình đã công bố gần đây cho thấy phương pháp tìm kiếm ảnh bằng kỹ thuật gom cụm là khả thi và đạt kết quả tốt.

Năm 2013, Hernández-Gracidias *et al.* thực hiện nâng cao hiệu suất tìm kiếm ảnh dựa trên quan hệ không gian 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 [14]. 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.

Năm 2017, Hakan *et al.* 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 [15]. Phương pháp đã thực nghiệm trên bộ ảnh ImageCLEF và so sánh tính hiệu quả với các phương pháp khác. Tuy nhiên, phương pháp này chưa phân tích ngữ nghĩa của hình ảnh cũng như không xây dựng cấu trúc dữ liệu nhằm tăng tốc độ tra cứu hình ảnh.

Năm 2017, Van *et al.* 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 [16]. Tuy nhiên, nhóm tác giả vẫn chưa trích xuất được ngữ nghĩa cho hình ảnh và tìm kiếm dựa trên ngữ nghĩa này.

Công cụ I2T (Image to Text) được Yao *et al.* giới thiệu năm 2010 nhằm tạo ra các RDF mô tả ngữ nghĩa hình ảnh để từ đó người dùng có thể truy vấn thông qua ngữ nghĩa này. Đồ

thị AoG (and-or Graph) được sử dụng để chuyển đổi quan hệ của các thành phần trong hình ảnh trở thành ngữ nghĩa tự nhiên nhằm mô tả hình ảnh [17]. Đây là một hệ thống tìm kiếm ngữ nghĩa của hình ảnh và giúp cho bài toán tra cứu ảnh theo tiếp cận ngữ nghĩa được hoàn thiện hơn.

Vijayarajan *et al.* thực hiện tra cứu ảnh dựa trên việc phân tích ngôn ngữ tự nhiên để tạo ra câu truy vấn SPARQL nhằm tìm kiếm tập hình ảnh dựa trên RDF-mô tả ảnh vào năm 2016 [12]. Quá trình tìm kiếm hình ảnh phụ thuộc vào việc phân tích văn phạm của ngôn ngữ để tạo thành các từ khóa mô tả nội dung hình ảnh. Phương pháp này chưa thực hiện phân lớp nội dung hình ảnh từ các đặc trưng màu sắc và đặc trưng không gian để tạo các từ khóa nhằm thực hiện tra cứu; do đó chưa thực hiện quá trình tìm kiếm từ một hình ảnh truy vấn cho trước.

Từ các công trình đã công bố cho thấy bài toán tra cứu ảnh có nhiều quan tâm của các nhóm tác giả. Trên cơ sở kế thừa từ 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 nhằm nâng cao hiệu quả tìm kiếm, nhóm tác giả đề xuất một mô hình truy vấn ảnh bằng cách tạo đồ thị đặc trưng không gian kết hợp lớp ngữ nghĩa và quan hệ không gian của hình ảnh; sau đó gom cụm các đồ thị này nhằm nâng cao hiệu suất cho quá trình tìm kiếm ảnh tương tự. Vì vậy, hệ tra cứu ảnh tương tự dựa trên đồ thị đặc trưng không gian được nhóm tác giả tiếp cận và thực hiện để từ đó nâng cao hiệu suất tìm kiếm ảnh.

### 3. PHƯƠNG PHÁP NGHIÊN CỨU

#### 3.1. Phân vùng ảnh và trích xuất đặc trưng

Mỗi hình ảnh trong tập dữ liệu được phân thành nhiều vùng khác nhau theo phương pháp của Hugo Jair Escalante [18]. Mỗi vùng được trích xuất một vector đặc trưng bao gồm đặc trưng thị giác và chú thích lớp ngữ nghĩa [19]. Bên cạnh đó, mỗi ảnh còn có một ma trận topo mô tả mối quan hệ không gian giữa các vùng trong ảnh. Hình 1 mô tả một ảnh gốc, 4 ảnh của các vùng thuộc về các lớp ngữ nghĩa: *sky*, *rock*, *trees*, *ground* và ma trận topo của ảnh 2968.jpg trong bộ dữ liệu ảnh ImageCLEF.



Hình 1. Ảnh gốc, các ảnh phân đoạn và ma trận topo của ảnh 2968.jpg trong tập dữ liệu

#### 3.2. Đồ thị đặc trưng không gian

**Định nghĩa 1.** Cho  $I_i, I_j$  lần lượt là hai vùng của ảnh  $I$ . Hai vùng  $I_i, I_j$  được gọi là kề nhau, ký hiệu  $topo(I_i, I_j) = 1$ , nếu giao điểm giữa chúng khác rỗng, ngược lại thì gọi là hai vùng tách rời, ký hiệu  $topo(I_i, I_j) = 2$ .

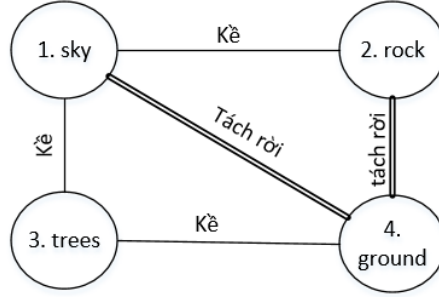
**Định nghĩa 2.** Đồ thị đặc trưng không gian là đồ thị vô hướng  $G = (V, E)$ , trong đó tập đỉnh  $V$ , tập cạnh  $E$  được định nghĩa như sau:

$$V = \{V_i | V_i = (id(I_i), l(I_i)), I_i \in I, i = 1..n\}$$

$$E = \{(V_i, V_j) | (V_i, V_j) \in \{1, 2\}; i, j = 1..n\}$$

Với  $I_i, I_j$  là vùng ảnh thứ  $i, j$  trong  $n$  vùng ảnh của  $I$ ;  $id(I_i), l(I_i)$  lần lượt là định danh và lớp ngữ nghĩa của vùng ảnh thứ  $i$ ;  $(V_i, V_j) = 1$  nếu hai vùng  $V_i$  và  $V_j$  kề nhau, ngược lại

(hai vùng  $V_i$  và  $V_j$  tách rời) thì  $(V_i, V_j) = 2$ . Hình 2 minh họa đồ thị đặc trưng không gian của ảnh 2968.jpg trong tập dữ liệu ảnh ImageCLEF.



Hình 2. Đồ thị đặc trưng không gian của ảnh 2968.jpg trong tập dữ liệu ImageCLEF

### 3.3. 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, độ đo tương tự giữa các đồ thị được đề xuất như sau:

**Định nghĩa 3:** (Độ tương tự theo lớp ngữ nghĩa giữa hai đồ thị)

$$S_l(G_I, G_J) = 1 - \frac{2 * f(G_C)}{f(G_I) + f(G_J)} \quad (1)$$

Với  $f(G_C)$  là số đỉnh chung của hai đồ thị;  $f(G_I)$ ,  $f(G_J)$  lần lượt là số đỉnh của đồ thị  $G_I$  và  $G_J$ .

**Mệnh đề 1:** Độ đo  $S_l(G_I, G_J)$  là một metric vì có các tính chất sau:

- (1) Không âm:  $S_l(G_I, G_J) \geq 0$  và  $S_l(G_I, G_J) = 0 \Leftrightarrow G_I = G_J$
- (2) Đối xứng:  $S_l(G_I, G_J) = S_l(G_J, G_I)$
- (3) Bất đẳng thức tam giác:  $S_l(G_I, G_J) + S_l(G_J, G_K) \geq S_l(G_I, G_K)$

**Chứng minh:** Dễ dàng chứng minh được  $S_l(G_I, G_J)$  thỏa 3 tính chất trên.

**Định nghĩa 4:** (Độ tương tự theo quan hệ không gian giữa hai đồ thị)

$$S_r(G_I, G_J) = 1 - \frac{2 * g(G_C)}{g(G_I) + g(G_J)} \quad (2)$$

Với  $g(G_C)$  là số cạnh chung của hai đồ thị theo quan hệ topo như ở **Định nghĩa 1**;  $g(G_I)$ ,  $g(G_J)$  lần lượt là số cạnh của đồ thị  $G_I$  và  $G_J$ .

**Mệnh đề 2.** Độ đo  $S_r(G_I, G_J)$  là một metric.

*Chứng minh:*

Theo **Mệnh đề 1** thì  $S_l(G_I, G_J)$  là một metric. Do đó dễ dàng kiểm chứng  $S_r(G_I, G_J)$  cũng là một metric.

Dựa trên công thức (1) và (2), độ đo tương tự giữa hai đồ thị đặc trưng không gian được tính theo công thức (3) như sau:

$$S(G_I, G_J) = \alpha S_l(G_I, G_J) + \beta S_r(G_I, G_J); \alpha + \beta = 1; \alpha, \beta \in (0,1) \quad (3)$$

**Mệnh đề 3.** Độ đo  $S(G_I, G_J)$  là một metric.

*Chứng minh:*

Theo **Mệnh đề 1** và **Mệnh đề 2** thì  $d_1(G_I, G_J)$  và  $d_2(G_I, G_J)$  là các metric. Do đó dễ dàng kiểm chứng được rằng  $d(G_I, G_J)$  cũng là một metric. ■

### 3.4. Thuật toán phân hoạch cụm đồ thị

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 cách 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, thuật toán K-means được cải tiến để áp dụng cho việc gom cụm đồ thị đặc trưng không gian 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 hai đồ thị trong quá trình gom cụm và cố định tâm cụm theo ngưỡng.

#### Thuật toán GC

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

**Đầu ra:** Tập cụm  $C$ .

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

    Khởi tạo  $C = \phi$ ;

**For**  $g \in G$  **do**

**If**  $C = \phi$  **then**

            Khởi tạo cụm  $C_0$  với tâm  $g$ ;

**Else**

            Tìm cụm  $C_k \in C : S(g, cen_k) = \min\{S(g, cen_i), i = 1, \dots, m\}$   
            // $cen_k$  là tâm của cụm  $k$

**If**  $S(g, cen_k) < \theta$  **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 4.** Độ phức tạp của **Thuật toán GC** là  $O(N^2)$ , với  $N$  là số đồ thị đặc trưng không gian trong tập  $G$ .

*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  $C_k \in C : S(g, centroid_k) = \min\{S(g, centroid_i), i = 1, \dots, m\}$ ” 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 GC** 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 GC** là  $O(N^2)$  ■

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

Trên cơ sở tập cụm kết quả ở pha tiền xử lý, thuật toán tìm tập ảnh tương tự được đề xuất. Với mỗi ảnh truy vấn, thuật toán thực hiện tìm tập ảnh tương tự. Quá trình này được thực hiện 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.

**Thuật toán IR**

**Đầu vào:** Đồ thị đặc trưng của ảnh truy vấn  $I_Q$ , tập cụm  $C$ , giá trị ngưỡng  $\theta$ .

**Đầu ra:** Tập ảnh tương tự  $IMG$ .

**Function**  $IR(I_Q, C, \theta)$

Khởi tạo  $IMG = \phi$ ;

Tim cụm  $C_k \in C: S(I_Q, cen_k) = \min \{S(I_Q, cen_i), i = 1, \dots, m\}$ ;

**If**  $S(I_Q, cen_k) < \theta$  **then**

$IMG = IMG \cup I_i, I_i \in C_k, i = 1, \dots, |C_k|$ ;

**End If**

Sắp xếp tập  $IMG$  giảm dần theo  $S(I_Q, I_i)$ ;

**Return**  $IMG$ ;

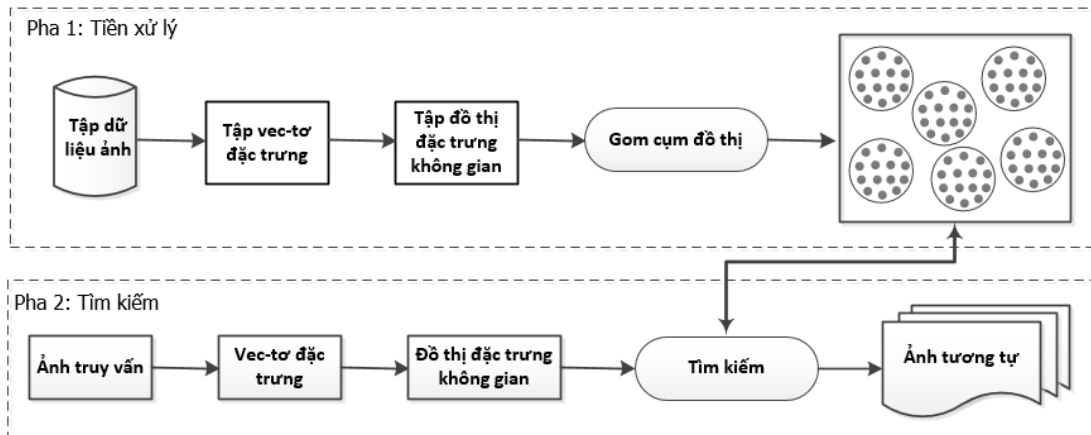
**End.**

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

*Chứng minh.* Gọi tập cụm  $C$  có số lượng cụm là  $n$ . 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 IR** là  $O(n)$ . ■

**4. KẾT QUẢ NGHIÊN CỨU**

**4.1. Mô hình thực nghiệm**



Hình 3. Mô hình hệ thống tìm kiếm ảnh dựa trên đồ thị đặc trưng không gian

Mô hình tổng quát của hệ thống tìm kiếm ảnh dựa trên đồ thị đặc trưng không gian 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 đặc trưng không gian từ tập lớp ngữ nghĩa và ma trận topo của 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 không gian 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.

#### 4.2. Kết quả thực nghiệm

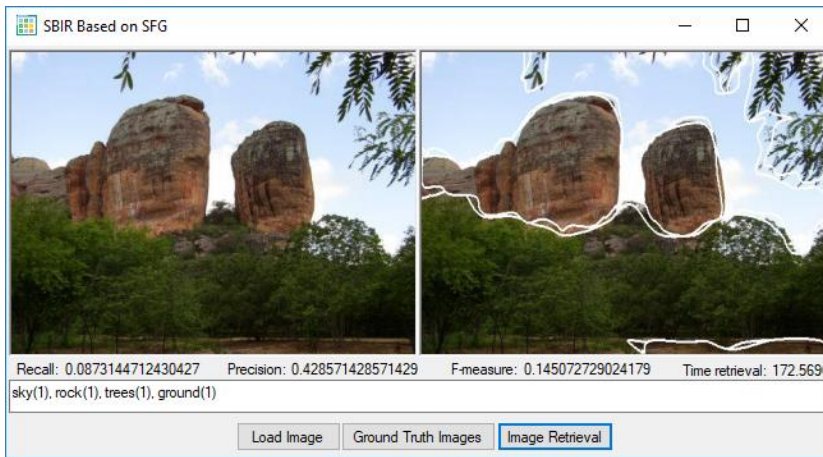
Thực nghiệm được thực thi trên máy PC CPU Intel Core i5-7200U CPU @ 2.50GHz, 8.0GB RAM, hệ điều hành Windows 10 Pro 64 bit. Pha tiền xử lý sử dụng ngôn ngữ lập trình Python phiên bản 3.6 và các gói thư viện *numpy*, *scipy.io*. Pha tìm kiếm ảnh sử dụng ngôn ngữ lập trình C# và .NET Framework 4.5. Các biểu đồ được tạo ra bằng ngôn ngữ Matlab. 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 có kích thước 1,64 GB. Để đánh giá hiệu quả của phương pháp tìm kiếm ảnh, phân 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 [20]:

$$precision = \frac{|relevant\ images \cap retrieved\ images|}{|retrieved\ images|}$$

$$recall = \frac{|relevant\ images \cap retrieved\ images|}{|relevant\ images|}$$

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

Trong đó, *relevant images* là tập ảnh tương tự với ảnh truy vấn 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].

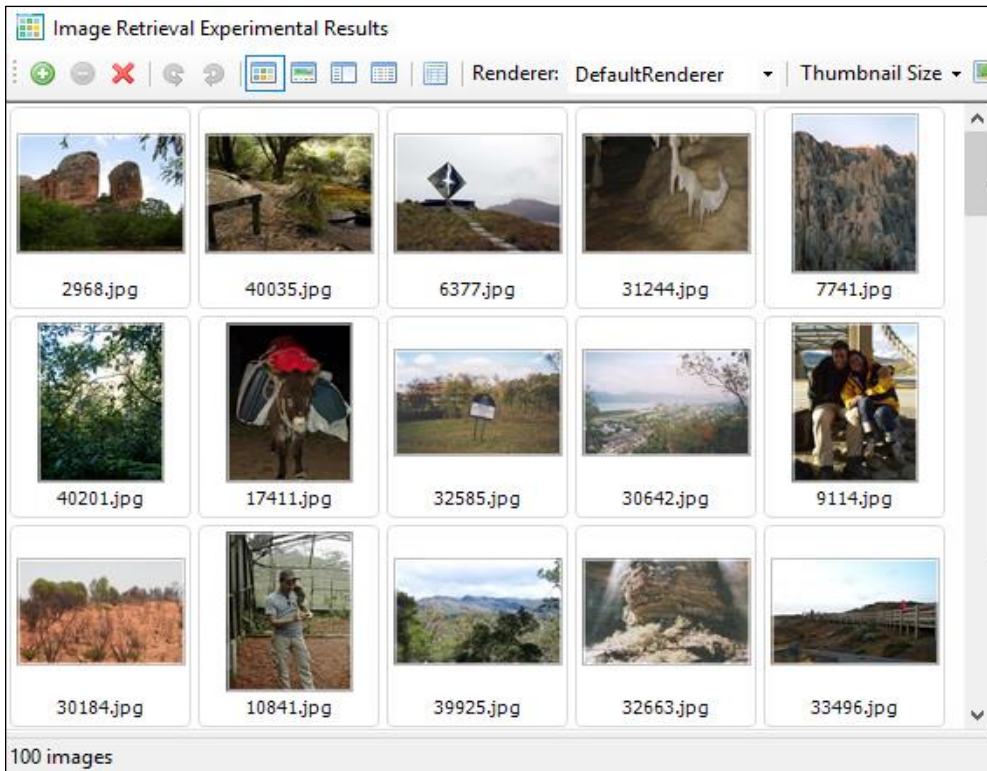


Hình 4. Ứng dụng thực nghiệm của phương pháp đề xuất

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

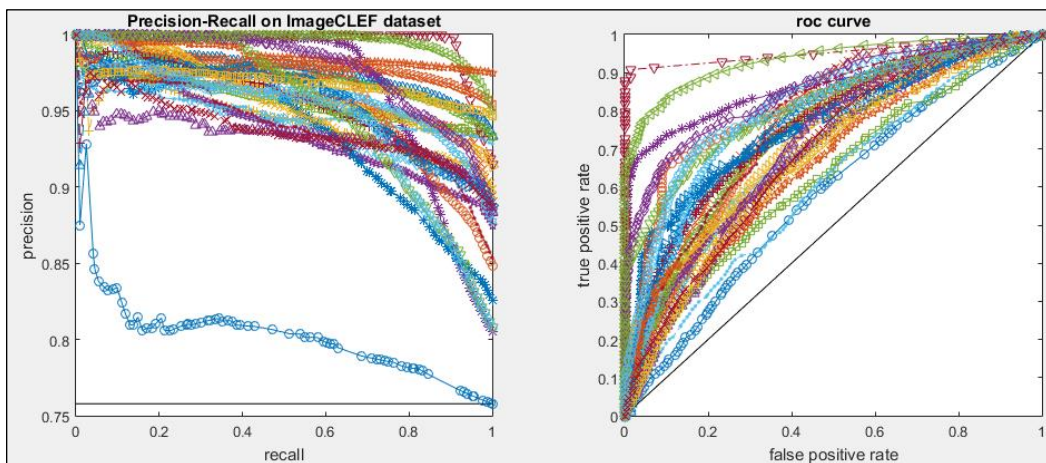
Tập ảnh	Số ảnh	Số cụm	Độ chính xác trung bình	Độ phủ trung bình	Độ đo dung hòa trung bình
00-10	6450	18	0,737182	0,424542	0,538794
11-20	4857	20	0,842857	0,389891	0,533154
21-30	3615	19	0,765874	0,402042	0,527287
31-40	5078	18	0,752721	0,463918	0,574042





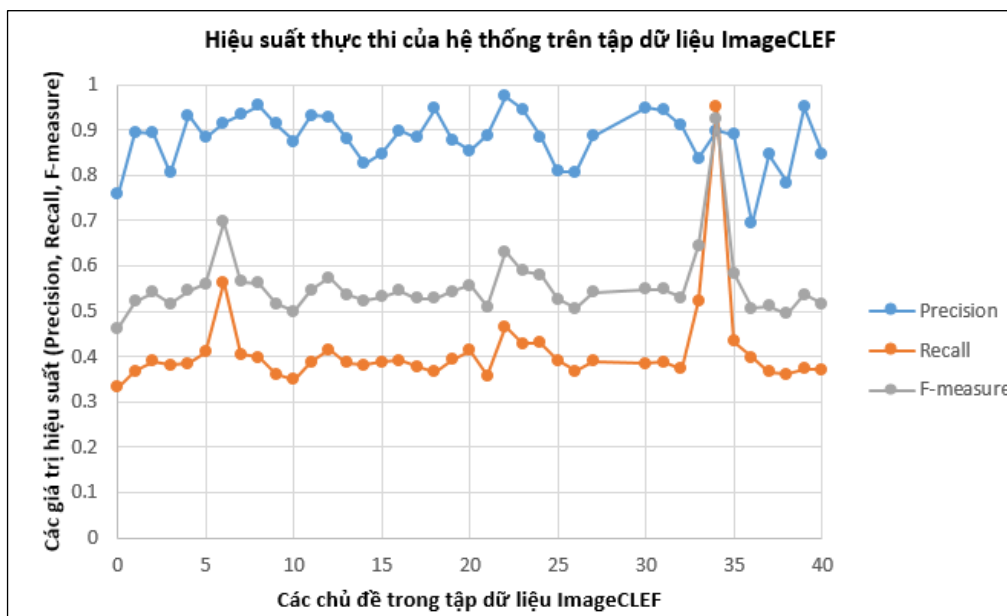
Hình 5. Kết quả tìm kiếm ảnh của phương pháp đề xuất

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 khác.



Hình 6. Precision-Recall và đường cong ROC của ứng dụng trên tập dữ liệu ImageCLEF





Hình 7. Trung bình Precision, Recall và F-measure 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 [15]	0,4678
C.A. Hernández-Gracidas, 2013 [14]	0,5826
Y. Cao, 2016 [21]	0,7236
Vijayarajan (2016) [12]	0,4618
Đề xuất của nhóm tác giả	0,7747

## 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo, nhóm tác giả đã xây dựng được phương pháp tìm kiếm ảnh dựa trên cấu trúc đồ thị đặc trưng không gian của hình ảnh. Dựa trên phương pháp, mô hình và các thuật toán đã đề xuất, hệ truy vấn ảnh được xây dựng và thực nghiệm trên bộ dữ liệu ảnh ImageCLEF. Kết quả thực được đánh giá dựa trên các giá trị recall, precision, F-measure và so sánh với các phương pháp khác trên cùng một tập dữ liệu ảnh đã cho thấy phương pháp đề xuất là hiệu quả. Vì vậy, phương pháp đề xuất của nghiên cứu này có thể được sử dụng cho các hệ thống tra cứu dữ liệu đa phương tiện. Hướng phát triển tiếp theo của bài báo là áp dụng kỹ thuật phân lớp để phân loại các phần tử trên cụm, đồng thời tạo ra câu truy vấn SPARQL từ vec-tơ từ và thực hiện truy vấn trên ontology để trích xuất ngữ nghĩa hình ảnh.

**Lời cảm ơn:** Nhóm tác giả chân thành cảm ơn Trường Đại học Công nghiệp Thực phẩm TP.HCM là nơi bảo trợ cho nghiên cứu này. Trân trọng cảm ơn nhóm nghiên cứu SBIR-HCM và Trường Đại học Sư phạm TP.HCM đã hỗ trợ về chuyên môn và cơ sở vật chất để nhóm tác giả hoàn thành nghiên cứu này.

## TÀI LIỆU THAM KHẢO

1. Guang-Hai Liu, Jing-Yu Yang. - Content-based image retrieval using color difference histogram, *Pattern Recognition* **46** (1) (2013) 188-198.
2. 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** (2003) 39-49.
3. Hun Woo Yoo, She Hwan Jung, Dong Sik Jang, Yoon Kyoona - Extraction of major object features using VQ clustering for content-based image retrieval, *Pattern Recognition* **35** (5) (2002) 1115-1126.
4. Deloitte - Photo sharing: trillions and rising, Deloitte Touche Tohmatsu Limited, Deloitte Global, Tech. Rep. (2016).
5. Xiao Xie, Xiwen Cai, Junpei Zhou, Nan Cao, Yingcai Wu - A semantic-based method for visualizing large image collections, *IEEE Transactions on Visualization and Computer Graphics*, 25 (7) (2018) 2362 - 2377.
6. 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.
7. 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.
8. Minakshi Banerjee, Sanghamitra Bandyopadhyay, Sankar K. Pal - A Clustering Approach to Image Retrieval Using Range Based Query and Mahalanobis Distance. in In: Skowron A., Suraj Z. (eds) *Rough sets and intelligent systems - Professor Zdzislaw Pawlak in Memoriam. 2013. Springer, Berlin, Heidelberg* (2013).
9. Mohammad Mehdi Saboorian, Mansour Jamzad, Hamid R. Rabiee - User adaptive clustering for large image databases. in *20th IEEE International Conference on Pattern Recognition. 2010. Istanbul, Turkey: IEEE* (2010) 4271-4274.
10. Zakariya S. M., Rashid Ali, Nesar Ahmad - Combining visual features of an image at different precision value of unsupervised content based image retrieval. in *2010 IEEE International Conference on Computational Intelligence and Computing Research. 2011. Coimbatore, India: IEEE* (2011).
11. Jabeen Safia, Mehmood Zahid, Mahmood Toqeer, Saba Tanzila, Rehman Amjad, Mahmood Muhammad Tariq - An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model, *PLoS ONE* **13** (4) (2018) 1-24.
12. Vijayarajan V., Dinakaran M., Tejaswin P., Lohani M. - A generic framework for ontology-based information retrieval and image retrieval in web data, *Human-centric Computing and Information Sciences* **6** (18) (2016) 1-30.
13. Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee, Song-Chun Zhu - I2T: Image Parsing to Text Description. in *Proceedings of the IEEE. 2010. IEEE* (2010) 1485-1508.
14. Hernández-Gracidas C.A., Sucar, L.E., Montes-y-Gómez M. - Improving image retrieval by using spatial relations, *Multimed Tools Application* **62** (2) (2013) 479-505.

15. Hakan Cevikalp Merve Elmas, Savas Ozkan - Large-scale image retrieval using transductive support vector machines, *Computer Vision and Image Understanding*, **173** (2018) 2-12.
16. Thanh The Van, Thanh Manh Le - Content-based image retrieval based on binary signatures cluster graph, *Journal of Knowledge Engineering, Expert System* **35** (1) (2017) 1-22.
17. Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee, Song-Chun Zhu - I2T: Image Parsing to Text Description. in *Proceedings of the IEEE* (2010) 1485-1508.
18. 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.
19. Carbonetto P.; Available from: <http://www.cs.ubc.ca/~pcarbo/>.
20. Ahmad Alzu'bi, Abbas Amira, Naeem Ramzan - Semantic content-based image retrieval: A comprehensive study, *Journal of Visual Communication and Image Representation* **32** (2015) 20-54.

## **ABSTRACT**

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

Nguyen Van Thinh\*, Dinh Thi Man  
Nguyen The Huu, Nguyen Thi Dinh, Van The Thanh  
*Ho Chi Minh City University of Food Industry*  
\*Email: [thinhnv@hufi.edu.vn](mailto:thinhnv@hufi.edu.vn)

In this paper, we build content-based image retrieval and improve efficient retrieval by proposing the spatial feature graph of the image. First of all, we propose a spatial feature graph of images and similarity measure between two images based on semantic label and spatial relations among regions in the image; Then, on the base of K-Means algorithm, several improvements to cluster the spatial feature graphs based on the similarity measure are proposed. Since then, we propose an image retrieval algorithm rely 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:* Image retrieval, clustering, spatial feature, similarity measure.