

MỘT PHƯƠNG PHÁP TRA CỨU ẢNH HIỆU QUẢ SỬ DỤNG PHÂN CỤM PHỔ TRONG PHẢN HỒI LIÊN QUAN

Đào Thị Thúy Quỳnh¹, Nguyễn Hữu Quỳnh², Phương Văn Cảnh², Ngô Quốc Tạo³

¹Trường Đại học Khoa học, Đại học Thái Nguyên, ² Khoa Công nghệ thông tin, Trường Đại học Điện lực,

³Viện Công nghệ thông tin, Viện Hàn Lâm Khoa học và Công nghệ Việt Nam,

quynhdt@tus.edu.vn, quynhnh@epu.edu.vn, canhpv@epu.edu.vn, ngtao@ioit.ac.vn

TÓM TẮT: Nhiều kỹ thuật tra cứu ảnh dựa vào nội dung được thiết kế để lấy ra các ảnh trong một lân cận nào đó của ảnh truy vấn và do đó bỏ qua các ảnh liên quan nằm trong toàn bộ không gian đặc trưng. Trong bài báo này, chúng tôi đề xuất một phương pháp tra cứu ảnh, gọi là **SCRF (spectral clustering in relevant feedback)** có ưu điểm là không yêu cầu người dùng phải xây dựng truy vấn phức tạp mà vẫn lấy được ảnh nằm rải rác trong toàn bộ không gian đặc trưng. Bên cạnh đó, phương pháp khai thác được đầy đủ thông tin tương tự giữa các ảnh phản hồi của người dùng hình thành các cụm liên quan ngữ nghĩa để xây dựng truy vấn đa điểm ở lân cận truy vấn tiếp theo. Hơn nữa, thời gian tra cứu của phương pháp cũng không tăng theo số lượng ảnh phản hồi từ người dùng. Chúng tôi cũng cung cấp các kết quả thực nghiệm để minh chứng độ chính xác của phương pháp.

Từ khóa: Tra cứu ảnh dựa vào nội dung, phản hồi liên quan, truy vấn đa điểm, phân cụm phổ.

I. GIỚI THIỆU

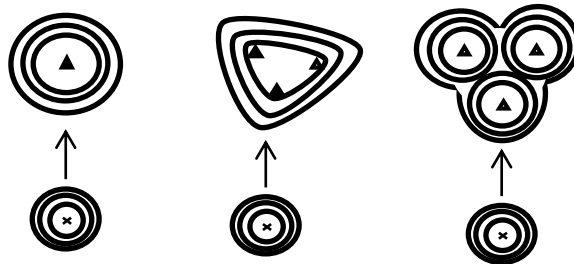
Tra cứu ảnh dựa vào nội dung (CBIR-Content Based Image Retrieval) đã nhận được nhiều sự quan tâm trong thập kỷ qua, do nhu cầu xử lý hiệu quả lượng dữ liệu đa phương tiện khổng lồ và tăng nhanh chóng. Nhiều hệ thống CBIR đã được phát triển, gồm QBIC, Photobook, MARS, NeTra, PicHunter, Blobworld, VisualSEEK, SIMPLicity và những hệ thống khác. Trong một hệ thống CBIR tiêu biểu, các đặc trưng ảnh trực quan mức thấp (tức là màu, kết cấu và hình dạng) được trích rút tự động cho mục tiêu đánh chỉ số và mô tả ảnh. Đối với cách tiếp cận truy vấn bởi mẫu, một ảnh truy vấn đưa vào hệ thống sẽ được xử lý tương tự như ảnh cơ sở dữ liệu để sinh ra một vectơ thích hợp. Tra cứu tiếp theo được thực hiện bằng việc sinh ra một danh sách các ảnh được phân hạng theo thứ tự giảm dần của độ đo tương tự so với ảnh truy vấn.

Là một vấn đề quan trọng trong CBIR, độ đo tương tự lượng hóa sự giống nhau về nội dung giữa từng cặp ảnh. Phụ thuộc vào kiểu đặc trưng mà chúng ta lựa chọn độ đo tương tự thích hợp. Tất cả các kỹ thuật tra cứu dựa vào nội dung hiện nay đều thừa nhận thông tin tương hỗ giữa độ đo tương tự ảnh và ngữ nghĩa của ảnh. Bằng các cách khác nhau, độ đo tương tự cố gắng nắm được một khía cạnh nào đó của nội dung ảnh, đó là ngữ nghĩa kế thừa từ độ tương tự hay đặc trưng mức thấp. Tuy nhiên, ngữ nghĩa kế thừa từ độ tương tự nhiều khi không giống với khái niệm mức cao được truyền tải bởi một ảnh (ngữ nghĩa của ảnh). Đó chính là khoảng cách ngữ nghĩa [7], nó phản ánh sự khác biệt giữa năng lực mô tả hạn chế của đặc trưng trực quan mức thấp và khái niệm mức cao. Cách tiếp cận dựa vào phản hồi liên quan đối với tra cứu ảnh dựa vào nội dung là một lĩnh vực nghiên cứu tích cực trong mấy năm qua nhằm rút ngắn khoảng cách ngữ nghĩa. Một số nghiên cứu tốt theo cách tiếp cận này có thể tìm thấy trong [1; 3; 8; 10; 11; 13; 14; 16]. Hầu hết các hệ thống CBIR đã có biểu diễn các ảnh bằng các vectơ đặc trưng sử dụng các đặc trưng trực quan, trong đó hai vectơ được coi là gần nhau nếu hai ảnh tương ứng với hai vectơ đó sẽ tương tự nhau hơn. Khi các hệ thống CBIR đưa ra một tập các ảnh được xem là tương tự với một ảnh truy vấn đã cho, người dùng có thể lấy ra các ảnh liên quan nhất đối với truy vấn đã cho và hệ thống điều chỉnh lại truy vấn sử dụng các ảnh liên quan mà người dùng vừa chọn. Các kỹ thuật CBIR dựa vào phản hồi liên quan không yêu cầu người dùng cung cấp các truy vấn khởi tạo chính xác nhưng yêu cầu người dùng xây dựng truy vấn lý tưởng thông qua đánh giá các ảnh là liên quan hay không.

Các cách tiếp cận đối với CBIR giả thiết rằng, về nguyên tắc các ảnh liên quan gần với ảnh truy vấn trong không gian đặc trưng nào đó. Tuy nhiên, sự tương tự giữa các ảnh mà con người nhận thức lại có sự khác biệt với khoảng cách giữa chúng trong không gian đặc trưng. Tức là, các ảnh liên quan về mặt ngữ nghĩa có thể nằm phân tán trong toàn bộ không gian đặc trưng và nằm rải rác ở một số cụm chứ không phải một cụm. Trong trường hợp này, các cách tiếp cận phản hồi liên quan truyền thống [1; 3; 5; 6; 8; 10; 11; 14; 16; 18; 19] không làm việc tốt khi dịch chuyển tâm truy vấn.

Thực hiện phản hồi liên quan đề cập đến việc tính toán một hoặc nhiều điểm truy vấn mới trong không gian đặc trưng và thay đổi hàm khoảng cách. Như được chỉ ra trong Hình 1(a), các nghiên cứu theo hướng tiếp cận ban đầu [1; 5; 8; 16] biểu diễn một truy vấn mới bằng một điểm đơn và thay đổi các trọng số của các thành phần đặc trưng để tìm một điểm truy vấn tối ưu và một hàm khoảng cách tối ưu. Trong trường hợp này, một điểm đơn được tính toán sử dụng trung bình trọng số của tất cả các ảnh liên quan trong không gian đặc trưng. Các đường viền biểu diễn các đường có độ tương tự tương đương. Trong khi đó, một cách tiếp cận nghiên cứu sau đó [7; 20; 21; 22; 24] biểu diễn một truy vấn mới bằng nhiều điểm để xác định hình của đường viền như Hình 1(b). Cách tiếp cận này sử dụng một phương pháp phân cụm [23] để tính toán các điểm truy vấn mới sử dụng các kết quả truy vấn (các ảnh liên quan) dựa vào đánh giá phản hồi của người dùng. Với giả thiết rằng các ảnh liên quan được ánh xạ sang các điểm gần nhau theo độ đo tương tự. Một đường viền rộng được xây dựng để phủ tất cả các điểm truy vấn và hệ thống tìm các ảnh tương tự với

các truy vấn này. Tuy nhiên, nếu không gian đặc trưng và hàm khoảng cách rất khác so với nhận thức của người dùng, các ảnh liên quan được ánh xạ sang các vùng có hình dạng bất kỳ tách rời trong không gian đặc trưng. Tức là, các ảnh liên quan có thể được phân hạng dưới các ảnh được tra cứu khác theo một truy vấn đã cho. Để hội tụ nhanh đến nhu cầu thông tin ở mức ngữ nghĩa cao hơn, hệ thống sẽ tìm các ảnh tương tự với bất kỳ các điểm truy vấn nào như trong Hình 1(c). Một truy vấn mà tra cứu các ảnh tương tự với bất kỳ các điểm truy vấn nào được gọi là truy vấn tách rời hay truy vấn đa điểm. Đặc biệt, một truy vấn ảnh phức tạp được biểu diễn bằng nhiều vùng tách rời do các ảnh liên quan ngữ nghĩa có thể nằm rải rác trong một số vùng trực quan hơn là một vùng.



Hình 1.1. Hình dạng truy vấn

(a) Dịch chuyển điểm truy vấn (b) Hình dạng lỗi (đa điểm) (c) Hình dạng lõm (đa điểm)

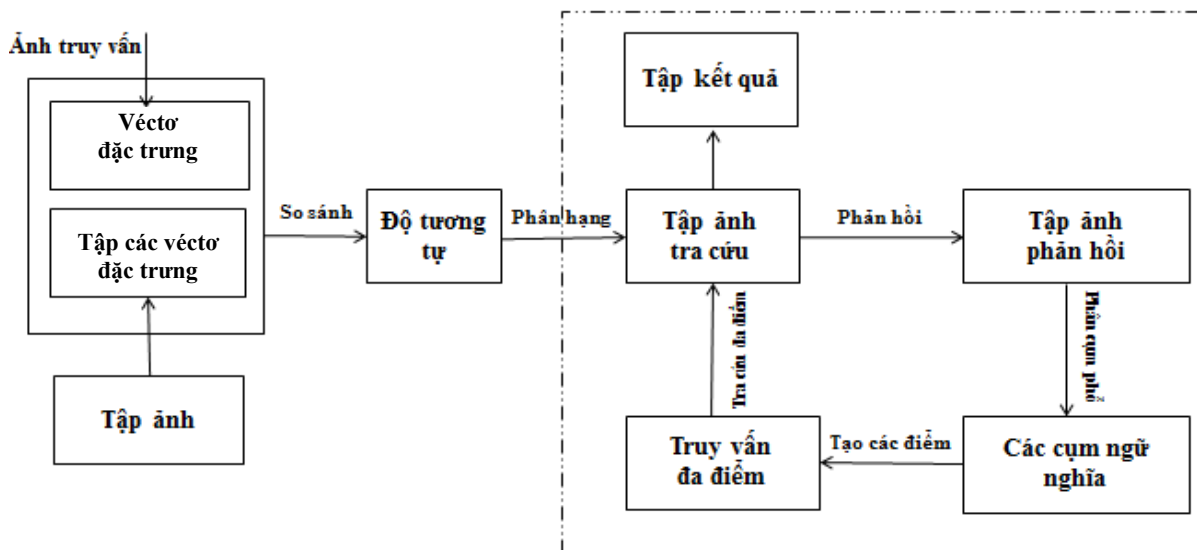
Tất cả các kỹ thuật CBIR hiện nay đều chắc chắn thừa nhận thông tin tương hỗ giữa độ đo tương tự và ngữ nghĩa của ảnh. Một hệ thống CBIR điển hình xếp hạng các ảnh mục tiêu theo độ đo tương tự đối với ảnh truy vấn nên chỉ lấy được các ảnh nằm trong lân cận của ảnh truy vấn và bỏ qua những ảnh liên quan nằm rải rác trong toàn bộ không gian đặc trưng. Các hạn chế ở trên là động lực để chúng tôi đề xuất phương pháp cải thiện được sự tương tác người dùng với các hệ thống tra cứu ảnh bằng cách khai thác đầy đủ thông tin độ tương tự giữa các ảnh trong tập phản hồi. Bên cạnh đó không cần đòi hỏi người dùng phải đưa vào nhiều ảnh truy vấn đa dạng thích hợp để biểu diễn nhu cầu thông tin của mình. Thời gian tra cứu cũng không tăng theo số lượng ảnh phản hồi của người dùng.

Phần còn lại của bài báo này được tổ chức như sau: trong phần 2, trình bày chi tiết phương pháp tra cứu ảnh sử dụng phần cụm phổ trong phản hồi liên quan. Phần 3, mô tả các kết quả thực nghiệm và cuối cùng là kết luận được đưa ra trong phần 4.

II. PHƯƠNG PHÁP TRA CỨU ẢNH HIỆU QUẢ SỬ DỤNG PHẦN CỤM PHỔ TRONG PHẢN HỒI LIÊN QUAN

Trong phần này, chúng tôi sẽ giới thiệu chung hệ thống đề xuất. Tiếp theo, chúng tôi mô tả chi tiết từng thành phần của hệ thống. Cuối cùng, thuật toán tra cứu đề xuất được trình bày.

2.1. Mô tả chung về phương pháp



Hình 2.1. Cấu trúc của phương pháp đề xuất

Phương pháp SCRF được mô tả bởi sơ đồ trên hình 2.1. Quá trình tra cứu bắt đầu từ việc trích rút đặc trưng của ảnh truy vấn. Các đặc trưng của ảnh cơ sở dữ liệu thường được trích rút và lưu trữ thành tập các vectơ đặc trưng. Sử dụng các đặc trưng này với một độ đo tương tự đặc trưng, sự tương đồng giữa ảnh truy vấn và ảnh cơ sở dữ liệu được so sánh và phân hạng. Tiếp theo, một tập ảnh lân cận với ảnh truy vấn khởi tạo được trả về cho người dùng. Người dùng sẽ chọn những ảnh liên quan tới mong muốn của họ để hình thành lên tập ảnh phản hồi. Một thuật toán phân cụm

sẽ được áp dụng lên tập ảnh phản hồi để hình thành lên các cụm liên quan ngữ nghĩa. Với mỗi cụm vừa tìm được phương pháp của chúng tôi sẽ thực hiện tìm đại diện cho mỗi cụm để hình thành truy vấn đa điểm đưa vào thực hiện tra cứu ở lần lặp sau. Quá trình được lặp lại cho đến khi người dùng ngừng phản hồi và phương pháp đưa ra tập kết quả.

2.2. Phương pháp đề xuất

Phương pháp của chúng tôi thay vì tìm một truy vấn trung tâm cho các mẫu tích cực mà người dùng chọn, chúng tôi sẽ thực hiện phân cụm tập ảnh phản hồi của người dùng. Sau khi có được các cụm ngữ nghĩa đó, chúng tôi tìm đại diện cho mỗi cụm. Mỗi đại diện đó được dùng để hình thành lên truy vấn đa điểm ở lần lặp tra cứu tiếp theo. Phương pháp sẽ tìm các ảnh tương tự với bất kỳ điểm nào hay đại diện nào của truy vấn đa điểm để trả về danh sách ảnh đa dạng nằm rải rác trong toàn bộ không gian đặc trưng.

Thuật toán phân cụm tập ảnh phản hồi từ người dùng

Trong tập ảnh lân cận được trả về bởi truy vấn khởi tạo người dùng sẽ chọn n ảnh liên quan. Để khai thác thông tin tương tự giữa các ảnh trong tập ảnh phản hồi chúng ta gọi thuật toán CRISE để hình thành lên các cụm ngữ nghĩa. Mỗi ảnh được chọn để đại diện cho mỗi cụm phải là ảnh mà tương tự nhất với tất cả các ảnh trong cụm. Các đại diện của các cụm sẽ hình thành lên truy vấn đa điểm ở lần lặp tra cứu tiếp theo. Quá trình trên được lặp lại cho đến khi người dùng dừng phản hồi.

Biểu diễn và phân cụm tập ảnh phản hồi

Dưới một biểu diễn đồ thị, phân cụm có thể được phát biểu tự nhiên như một bài toán phân hoạch đồ thị. Trong số nhiều phương pháp phân hoạch đồ thị phổ [4; 15; 9; 17] đã được áp dụng thành công với nhiều lĩnh vực trong thị giác máy tính gồm phân tích chuyển động [5], phân đoạn ảnh [9; 17] và nhận dạng đối tượng [15]. Trong bài báo này, chúng tôi sử dụng phương pháp sử dụng k véctơ riêng và tính trực tiếp phân hoạch k -way trong [2]. So với phương pháp sử dụng một véctơ riêng tại một thời điểm và gọi đệ quy [9], phương pháp sử dụng k véctơ riêng được chỉ ra là tốt hơn về mặt thực hành. Nói chung, một phương pháp phân hoạch đồ thị cố gắng tổ chức các nút thành các nhóm sao cho độ tương tự trong phạm vi nhóm là cao, và/hoặc độ tương tự giữa các nhóm là thấp. Một đồ thị đã cho $G=(V,E)$ với ma trận affinity A , một cách đơn giản để định lượng giá cho các nút phân hoạch thành hai tập rời nhau C_1 và C_2 ($C_1 \cap C_2 = \emptyset$ và $C_1 \cup C_2 = V$) là tổng có trọng số của các cạnh mà kết nối hai tập. Tiếp theo, chúng tôi trình bày ngắn gọn phương pháp dựa trên nghiên cứu của A. Y. Ng và cộng sự (xem chi tiết hơn tại [2]).

Đầu tiên, từ n điểm dữ liệu ảnh, phương pháp xây dựng ma trận affinity A theo $a_{ij} = e^{-\frac{\|s_i - s_j\|^2}{2\sigma^2}}$ ($i \neq j$), $a_{ii}=0$ (1). Ở đây tham số tỉ lệ σ^2 điều khiển mức độ ái lực a_{ij} giảm nhanh thế nào với khoảng cách giữa s_i và s_j , phương pháp chọn tự động có thể xem trong [2]. Một giá trị a_{ij} giữa hai ảnh là “cao” nếu hai ảnh là rất tương tự.

Xây dựng ma trận đường chéo D trong đó phần tử (i,i) là tổng hàng thứ i của ma trận A . D là một ma trận chéo với $D_{ii} = \sum_{j=1, \dots, n} a_{ij}$.

Tính ma trận Laplace chuẩn hóa : $L = D^{-1/2} A D^{-1/2}$.

Tìm k véctơ riêng x_1, x_2, \dots, x_k lớn nhất của ma trận L , trong đó $x_1=(x_{11}, x_{12}, x_{13}, \dots, x_{1n})$, $x_2=(x_{21}, x_{22}, x_{23}, \dots, x_{2n})$, $\dots, x_k=(x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn})$ và xây dựng ma trận $X = [x_1^T, x_2^T, \dots, x_k^T] \in \mathbb{R}^{n \times k}$, cụ thể:

	x_1^T	x_2^T	x_3^T	...	x_k^T
x_1	x_{11}	x_{21}	x_{31}	...	x_{k1}
x_2	x_{12}	x_{22}	x_{32}	...	x_{k2}
x_3	x_{13}	x_{23}	x_{33}	...	x_{k3}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_n	x_{1n}	x_{2n}	x_{3n}		x_{kn}

Xây dựng ma trận Y từ X bằng việc chuẩn hóa mỗi dòng của X là chiều dài đơn vị của ma trận Y ($Y_{ij} = \frac{x_{ij}}{(\sum_j x_{ij}^2)^{\frac{1}{2}}}$)

y_1	y_{11}	y_{12}	y_{13}	...	y_{1k}
y_2	y_{21}	y_{22}	y_{23}	...	y_{2k}
y_3	y_{31}	y_{32}	y_{33}	...	y_{3k}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_k	y_{n1}	y_{n2}			y_{nk}

Mỗi dòng của ma trận Y được xem như là một điểm trong không gian véctơ k chiều. Đến đây, sẽ có n điểm trong không gian R^k , phân cụm $(y_i)_{i=1..n}$ trong không gian R^k thành k cụm C_1, C_2, \dots, C_k thông qua K-Means. Cuối cùng, gán điểm s_i tới cụm j nếu và chỉ nếu hàng thứ i của ma trận Y tương ứng với cụm j .

Hình 2 dưới đây là thuật toán phân cụm sử dụng k véctơ riêng **CRISE (Clustering Relevant Images Set using Eigenvectors)** thực hiện việc phân cụm tập các ảnh liên quan mà người dùng chọn thành k cụm.

Thuật toán CRISE

Input: - Tập các ảnh $S = \{s_1, s_2, \dots, s_n\}$ với $s_i \in R^n$
 - Số cụm k

Output: k cụm: C_1, C_2, \dots, C_k

Bước 1: Xây dựng ma trận affinity
 for $i \leftarrow 1$ to n do
 for $j \leftarrow 1$ to n do
 if ($i \neq j$) $a_{ij} \leftarrow \exp\left(\frac{-\|s_i - s_j\|^2}{2\sigma^2}\right)$
 else $a_{ij} \leftarrow 0$

Bước 2: Xây dựng ma trận đường chéo và ma trận Laplace L
 for $i \leftarrow 1$ to n do
 $d_{ii} \leftarrow \sum_{j=1..n} a_{ij}$
 $L \leftarrow D^{-1/2} A D^{-1/2}$

Bước 3: Tìm k véctơ riêng lớn nhất x_1, x_2, \dots, x_k của ma trận Laplace L
 for $i \leftarrow 1$ to k do
 $x_i \leftarrow \text{Largest_eigen_vectors}(L)$
 $X \leftarrow [x_1^T, x_2^T, \dots, x_k^T]$

Bước 4: Xây dựng ma trận Y từ X
 for $i \leftarrow 1$ to n do
 for $j \leftarrow 1$ to k do
 $y_{ij} \leftarrow x_{ij} / (\sum_k x_{ik}^2)^{1/2}$
 $Y \leftarrow [y_1, y_2, \dots, y_k]$

Bước 5: Phân thành k cụm thông qua K-Means
 $P \leftarrow \emptyset$
 for $i \leftarrow 1$ to n do
 $p_i \leftarrow y_i$
 $P \leftarrow P \cup p_i$
 K-Mean(P)

Bước 6: Gán các s_i vào các cụm
 for $i \leftarrow 1$ to n do
 if $p_i \in (C_j)_{j=1..k}$
 $C_j \leftarrow C_j \cup s_i$

Return C_1, C_2, \dots, C_k

Hình 2.2. Thuật toán CRISE

Tìm ảnh đại diện cho cụm

Để thực hiện việc tra cứu ảnh hiệu quả, một ảnh đại diện thích hợp phải thu được cho mỗi cụm. Ở đây, một ảnh được chọn là đại diện cho một cụm phải là ảnh mà tương tự nhất với tất cả các ảnh trong cụm. Phát biểu này được minh họa bằng toán học như sau: Với một biểu diễn đồ thị của các ảnh được cho $G=(V,E)$ với ma trận affinity A , cho tập các cụm ảnh là $\{C_1, C_2, \dots, C_k\}$ (tập các cụm này cũng này cũng là một phân hoạch của V , tức là $C_i \cap C_j = \emptyset, i \neq j$ và $\bigcup_{i=1}^k C_i = V$) thì ảnh đại diện của C_i là

$$\arg \max_{j \in C_i} \sum_{t \in C_i} a_{jt} \quad (2)$$

Như vậy, với một cụm, ảnh đại diện là ảnh mà có tổng độ tương tự trong phạm vi cụm là cực đại.

Khoảng cách từ một ảnh đến truy vấn đa điểm

Khác với các phương pháp tra cứu ảnh khác, phương pháp của chúng tôi sẽ hình thành lên truy vấn đa điểm $MQ=(Q_1, Q_2, \dots, Q_k)$ từ các đại diện của mỗi cụm. Khi đó, khoảng cách từ một ảnh DI_i đến truy vấn đa điểm $MQ=(Q_1, Q_2, \dots, Q_k)$ là cực tiểu của các khoảng cách có trọng số từ một ảnh DI_i đến mỗi Q_j trong truy vấn đa điểm và được tính theo công thức (3):

$$D(DI_i, MQ) = \min_{j=1..k} \text{dist}(DI_i, Q_j) \quad (3)$$

Trong công thức (3), $\text{dist}(DI_i, Q_j)$ với $i=1 \dots, N, j=1 \dots, k$ là khoảng cách từ một ảnh DI_i đến một điểm truy vấn Q_j trong truy vấn đa điểm MQ .

Tập tin cây nền (ground truth):

Tập tin cây nền Corel được sử dụng rộng rãi trong đánh giá CBIR, do đó chúng tôi cũng sử dụng phân loại Corel làm tin cây nền, tức là chúng tôi xem tất cả các ảnh trong cùng loại Corel là liên quan. Tập tin cây nền này gồm 3 cột (có tiêu đề: ID ảnh truy vấn, ID ảnh và Sự liên quan) và gồm 1,981,320 dòng.

3.2. Chiến lược mô phỏng phản hồi liên quan

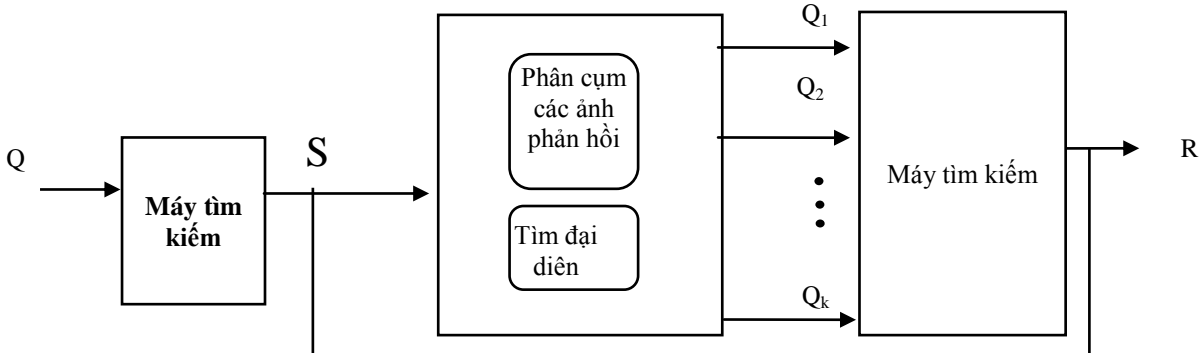
Để bắt chước hành vi của con người, chúng tôi thực hiện mô phỏng phản hồi liên quan trong thử nghiệm. Đầu tiên, truy vấn khởi tạo sẽ được thực hiện để tạo ra kết quả truy vấn. Chúng tôi mô phỏng tương tác người dùng bằng việc chọn n ảnh liên quan từ kết quả tra cứu khởi tạo dựa vào tập tin cây nền (ground truth). Những ảnh liên quan từ lần lặp phản hồi đầu tiên sẽ được phân thành k cụm và thực hiện tìm đại diện cho k cụm này. Sau đó k đại diện được dùng để xây dựng truy vấn đa điểm phục vụ cho tra cứu tiếp theo. Sau đó những kết quả tra cứu được gộp lại để tạo ra một danh sách kết quả tổng hợp theo chiến lược truy vấn đa điểm tách rời.

Phản hồi liên quan được thực hiện theo chiến lược chọn những ảnh liên quan đầu tiên (dựa vào tập tin cây nền) trong danh sách kết quả. Trong chiến lược này, trường hợp xấu nhất là không có ảnh liên quan nào ngoài ảnh truy vấn và trường hợp tốt nhất là có $n-1$ ảnh liên quan ngoài ảnh truy vấn. Do đó, số lượng ảnh liên quan có thể dao động từ 1 đến n ảnh (bao gồm cả ảnh truy vấn). Chiến lược này được sử dụng để mô phỏng người dùng thực tế trong thực nghiệm của chúng tôi.

3.3. Thực hiện truy vấn và đánh giá

Trong thực nghiệm của chúng tôi, các yếu tố đó được lựa chọn như sau:

Một truy vấn khởi tạo được đưa vào hệ thống, kết quả tương ứng với truy vấn đó được hiển thị cho người dùng. Sau đó, người dùng sẽ phản hồi trên danh sách kết quả tương ứng với truy vấn khởi tạo để hình thành danh sách ảnh phản hồi. Hệ thống sẽ thực hiện phân cụm danh sách ảnh phản hồi và tìm đại diện cho mỗi cụm. Đại diện của mỗi cụm sẽ xây dựng lên truy vấn đa điểm ở lần lặp truy vấn tiếp theo. Trong pha tính khoảng cách, khoảng cách từ một ảnh trong cơ sở dữ liệu đến truy vấn đa điểm là giá trị cực tiểu của các khoảng cách từ ảnh cơ sở dữ liệu tới một đại diện của truy vấn đa điểm để lấy được các ảnh nằm rải rác trong toàn bộ không gian đặc trưng. Quá trình sẽ dừng lại khi người dùng không tiếp tục phản hồi. Mô hình hệ thống thực hiện quá trình này được thể hiện trên Hình 3.3.



Hình 3.3. Mô hình hệ thống

Độ chính xác² trung bình ở mức 100 ảnh trả về được sử dụng để đánh giá. Bốn thiết lập phản hồi được sử dụng để so sánh là 1, 2, 3, 4 số đại diện của một truy vấn đa điểm và một chiến lược phản hồi, do đó sẽ có 4 cấu hình. Ba phương pháp khác nhau được sử dụng để so sánh bao gồm **CCH** (hệ thống CBIR truyền thống với một biểu diễn C+ trên đặc trưng toàn cục), hệ thống **ERIN** [12] với hệ thống **SCRF** mà chúng tôi đề xuất.

Bảng 2. Bảng kết quả của 3 phương pháp số đại diện của truy vấn đa điểm trong một lần phản hồi

Phương pháp	Độ chính xác theo số đại diện của truy vấn đa điểm			
	1	2	3	4
CCH	0.20	0.22	0.23	0.24
ERIN	0.24	0.29	0.31	0.33
SCRF	0.35168	0.43178	0.48154	0.48278

Trong Bảng 2, thể hiện độ chính xác trung bình của ba phương pháp là **CCH**, **ERIN** và phương pháp **SCRF** của chúng tôi tại các mức 1, 2, 4, 6, 8, 10 số đại diện của truy vấn đa điểm với phương pháp của chúng tôi số cụm cũng chính là số truy vấn.

² Độ chính xác là tỉ số giữa số các ảnh liên quan với ảnh truy vấn trong tập kết quả trả về trên tổng số các ảnh trả về.

IV. KẾT LUẬN

Chúng tôi đã tập trung vào đề xuất phương pháp, có tên là **SCRF**, giải quyết hai vấn đề chính đó là: (1) tìm các ảnh liên quan ngữ nghĩa nằm rải rác trong toàn bộ không gian đặc trưng với độ chính xác cao và (2) thời gian tra cứu không tăng theo số phản hồi của người dùng. Để giải quyết được hai vấn đề này, chúng tôi đã tận dụng sự đánh giá của người dùng để hình thành tập ảnh liên quan và phân cụm chúng thành các cụm ngữ nghĩa nằm rải rác trong toàn bộ không gian đặc trưng và đại diện của mỗi cụm hình thành lên truy vấn đa điểm. Phương pháp sử dụng một thuật toán phân cụm phổ có ưu điểm phân cụm các ảnh được kết nối với nhau nhưng không nhất thiết phải nhóm vào trong một đường bao lỗi nên thực hiện tốt hơn các thuật toán phân cụm truyền thống. Từ đó có thể tra cứu được các ảnh nằm rải rác trong toàn bộ không gian đặc trưng và nâng cao độ chính xác.

Kết quả thực nghiệm của chúng tôi trên cơ sở dữ liệu đặc trưng gồm 10.800 ảnh đã chỉ ra rằng phương pháp được đề xuất **SCRF** cung cấp một độ chính xác cao hơn hẳn so với các phương pháp **RIBR** và phương pháp **ERIN**.

Chúng tôi xin chân thành cảm ơn đề tài: “Nghiên cứu phương pháp tra cứu ảnh dựa vào đa truy vấn”, mã số PTNTDD17.04 đã hỗ trợ.

TÀI LIỆU THAM KHẢO

- [1] Andre B, Vercauteren T, Buchner AM, Wallace MB, Ayache N (2012). Learning semantic and visual similarity for endomicroscopy video retrieval. *IEEE Transactions on Medical Imaging*. 31(6):1276-88.
- [2] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and algorithm. In *Proceedings Of Neural Information Processing Systems (NIPS)*, 2001.
- [3] A. W. M. Smeulders, M. Worring, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (12) (2000) 1349-1380.
- [4] Bartolini, I., Ciacci, P., Waas, F., (2001). Feedbackbypass: A new approach to interactive similarity query processing. In: *Proceedings of the 27th VLDB Conference*, Roma, Italy, pp. 201-210.
- [5] J. Costeira and T. Kanade, “A multibody factorization method for motion analysis,” in *Proc. Int. Conf. Computer Vision*, 1995, pp. 1071-1076.
- [6] D. Liu, K. A. Hua, K. Vu, and N. Yu, (2009) “Fast Query Point Movement Techniques for Large CBIR Systems”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, No. 5, pp. 729-743.
- [7] K. A. Hua, N. Yu, and D. Liu (2006). Query Decomposition: A Multiple Neighborhood Approach to Relevance Feedback Processing in Content-based Image Retrieval. In *Proceedings of the IEEE ICDE Conference*.
- [8] Ishikawa, Y., Subramanya, R., Faloutsos, C., (1998). Mind Reader: Querying databases through multiple examples. In: *Proceedings of the 24th VLDB Conference*, New York, USA, pp. 218-227.
- [9] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888-905, Aug. 2000.
- [10] Norton, D.; Heath, D.; and Ventura, D. (2016). Annotating images with emotional adjectives using features that summarize local interest points. *IEEE Transactions on Affective Computing*, Under Review.
- [11] M. Ortega-Binderberger and S. Mehrotra (2004). Relevance feedback techniques in the MARS image retrieval systems. *Multimedia Systems*, 9(6):535-547.
- [12] Quỳnh N. H., Quỳnh D. T. T., Tao N. Q., Dung C. V., Canh P. V., Sơn A. H. (2016). Một phương pháp tra cứu ảnh biểu diễn nhu cầu thông tin người dùng hiệu quả, Kỷ yếu hội nghị Quốc gia lần thứ 9 về Nghiên cứu cơ bản và ứng dụng trong Công nghệ thông tin (FAIR).
- [13] Rui, Y., Huang, T., Ortega, M., Mehrotra, S., (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* 8 (5), pp. 644-655.
- [14] Rui, Y., Huang, T., Chang, S. F., (1999). Image Retrieval: current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation* 10, 39-62.
- [15] S. Sarkar and P. Soundararajan, “Supervised learning of large perceptual organization: graph spectral partitioning and learning automata,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 5, pp. 504-525, May 2000.
- [16] T. Gevers and A. Smeulders (2004). Content-based image retrieval: An overview. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*. Prentice Hall.
- [17] Y. Weiss, “Segmentation using eigenvectors: a unifying view,” in *Proc. Int. Conf. Computer Vision*, 1999, pp. 975-982.
- [18] Flickner, M., Sawhney, H., Niblack, W., et al., (1995). Query by image and video content: The QBIC system. *IEEE Computer Magazine* 28 (9), 23-32.
- [19] Rocchio, J. J., (1971). Relevance feedback in information retrieval. In: Salton, G. (Ed.), *The SMART Retrieval System-Experiments in Automatic Document Processing*. Prentice Hall, Englewood Cliffs, NJ, pp. 313-323.

- [20] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman (2007). Total recall: Automatic query expansion with a generative feature model for object retrieval. In Proc. ICCV.
- [21] Porkaew, K., Chakrabarti, K., (1999). Query refinement for multimedia similarity retrieval in MARS. In: Proceedings of the 7th ACM Multimedia Conference, Orlando, Florida, pp. 235-238.
- [22] R. Arandjelović and A. Zisserman (2012). Three things everyone should know to improve object retrieval. In Proc. CVPR.
- [23] Charikar, M., Chekuri, C., Feder, T., Motwani, R., (1997). Incremental clustering and dynamic information retrieval. In: Proceedings of the ACM STOC Conference, pp. 626-635.
- [24] Quynh Dao Thi Thuy, Quynh Nguyen Huu, Canh Phuong Van, Tao Ngo Quoc (2017), An efficient semantic - Related image retrieval method, Expert Systems with Applications, Volume 72, pp. 30-41

AN EFFICIENT IMAGE RETRIEVAL METHOD USING SPECTRAL CLUSTERING IN RELEVANT FEEDBACK

Nguyen Huu Quynh, Dao Thi Thuy Quynh, Phương Văn Canh, Ngo Quoc Tao

ABSTRACT: Many previous techniques were designed to retrieve images in a certain neighborhood of the query image, thus bypassing the related images in the whole feature space. Besides, some designed techniques only care about similarity between query image and data image that neglects similarities among data images. In this paper, we propose an efficient image retrieval method using spectral clustering in relevant feedback (SCRF) which has advantages that do not require the user to provide initial queries correctly but also retrieve relevant images in the entire feature space. In addition, our method fully exploit the similarity information of feedback image and contrast multipoints query in next query. Furthermore, the retrieval time of our method also is not increase with the number of user feedback. We also provide experimental results to demonstrate the effectiveness of our method.

Keywords: Content based image retrieval, relevant feedback, multipoints query, spectral clustering.