Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ IX "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR'9)"; Cần Thơ, ngày 4-5/8/2016
DOI: 10.15625/vap.2016.00054

MỘT PHƯƠNG PHÁP TRA CỨU ẢNH BIỂU DIỄN NHU CẦU THÔNG TIN NGƯỜI DÙNG HIỆU QUẢ

Nguyễn Hữu Quỳnh[1], Đào Thị Thúy Quỳnh[2], Ngô Quốc Tạo[3], Cù Việt Dũng[1], Phương Văn Cảnh[1], An Hồng Sơn[4]

[1]Khoa Công nghệ thông tin, Trường Đại học Điện lực,
[2]Trường Đại học Khoa học, Đại học Thái Nguyên,
[3]Viện Công nghệ thông tin, Viện Hàn Lâm Khoa học và Công nghệ Việt Nam,
[4]Trường Đại học Công nghiệp Việt Hung

quynhnh@epu.edu.vn, quynhdtt@tnus.edu.vn, nqtao@ioit.ac.vn, dungcv@epu.edu.vn, canhpv@epu.edu.vn, sonanhongvh@gmail.com

TÓM TẮT— *Hầu hết các cách tiếp cận tra cứu ảnh dựa vào nội dung truyền thống không biểu diễn hiệu quả nhu cầu thông tin của người dùng. Lý do của các hạn chế này là: (a) nhu cầu thông tin của người dùng rất phong phú, do đó khó có thể biểu diễn nhu cầu này với một ảnh truy vấn, (b) một ảnh thường gồm nhiều biểu diễn với độ quan trọng khác nhau nhưng các phương pháp thường coi độ quan trọng này là ngang nhau, (c) các đặc trưng mức thấp không phản ánh được thông tin ngữ nghĩa của ảnh và (d) hàm khoảng cách kết hợp với các đặc trưng mức thấp không thể hiện được nhận thức về độ tương tự trực quan của người dùng. Nhằm khắc phục hạn chế ở trên, chúng tôi đề xuất phương pháp tra cứu ảnh, có tên ERIN (Efficient Representation of Information Need). Phương pháp có ưu điểm biểu diễn tốt nhu cầu thông tin của người dùng do sử dụng nhiều ảnh và nhiều biểu diễn. Bên cạnh đó, phương pháp xác định được độ quan trọng của mỗi biểu diễn ảnh và giảm khoảng cách ngữ nghĩa giữa đặc trưng mức thấp và khái niệm mức cao dẫn đến nâng cao chất lượng hệ thống tra cứu ảnh. Chúng tôi đã thực nghiệm trên cơ sở dữ liệu ảnh gồm 10.800 ảnh. Các kết quả thực nghiệm chỉ ra rằng kỹ thuật này cải tiến được hiệu năng của hệ thống tra cứu ảnh dựa vào nội dung so với phương pháp đã có và cho kết quả gần với nhu cầu của người dùng.*

Từ khóa— *Tra cứu ảnh dựa vào nội dung, biểu diễn nhu cầu thông tin, đa truy vấn, véc tơ đặc trưng.*

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 [19], Photobook [4], MARS [25] NeTra [23], PicHunter [18] , Blobworld [6], VisualSEEK [28], SIMPLIcity [22] và những hệ thống khác [15, 32, 17, 16, 20, 24, 26, 21]. 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. Để tìm kiếm các ảnh mong muốn, người dùng đưa một ảnh làm mẫu và hệ thống trả lại một tập các ảnh tương tự dựa vào các đặc trưng được trích rút.

Cho dù nhiều thuật toán phức tạp đã được thiết kế để mô tả các đặc trưng màu, hình dạng và kết cấu, các thuật toán này không thể mô hình tương đương các ngữ nghĩa ảnh và có nhiều giới hạn khi giải quyết các cơ sở dữ liệu ảnh nội dung rộng [2]. Các thực nghiệm mở rộng trên các hệ thống CBIR chỉ ra rằng các nội dung mức thấp thường thất bại trong mô tả các khái niệm ngữ nghĩa mức cao trong ý nghĩa của người dùng [3]. Do đó, hiệu năng của CBIR vẫn còn xa so với các kỳ vọng của người dùng.

Trong [34], Eakins đã đề cập ba mức truy vấn trong CBIR, cụ thể: Mức 1: Tra cứu bởi các đặc trưng gốc như màu, kết cấu, hình dạng hoặc vị trí không gian của các thành phần ảnh. Truy vấn tiêu biểu là truy vấn bởi mẫu, "tìm những bức ảnh như cái này"; Mức 2: Tra cứu các đối tượng có loại đã cho được nhận biết bởi các đặc trưng gốc, với độ suy diễn logic nào đó. Chẳng hạn, "tìm bức ảnh có chứa một bông hoa hồng"; Mức 3: Tra cứu bởi các thuộc tính tóm tắt, bao gồm một lượng đáng kể lập luận mức cao về mục đích của các đối tượng hoặc các cảnh được miêu tả. Điều này bao gồm tra cứu của các sự kiện đã đặt tên, của các ảnh với xúc cảm hoặc tôn giáo,.. Truy vấn bởi mẫu, "tìm các ảnh của một đám đông vui nhộn". Mức 2 và 3 cùng nhau được gọi là tra cứu ảnh ngữ nghĩa và khoảng trống giữa các mức 1 và 2 là khoảng cách ngữ nghĩa [1]. Cụ thể hơn, sự khác nhau giữa khả năng mô tả của các đặc trưng ảnh mức thấp bị giới hạn và sự phong phú của ngữ nghĩa người dùng được gọi là khoảng cách ngữ nghĩa [5,27,35].

Các kỹ thuật trong việc rút ngắn "khoảng cách ngữ nghĩa" gồm có 5 loại chính: (1) sử dụng bản thể đối tượng để xác định các khái niệm mức cao, (2) sử dụng các công cụ học máy để kết hợp các đặc trưng mức thấp với các khái niệm truy vấn, (3) đưa phản hồi liên quan vào lặp tra cứu cho học ý định của người dùng, (4) sinh ra mẫu ngữ nghĩa để hỗ trợ tra cứu ảnh mức cao, (5) Cách sử dụng cả nội dung trực quan của các ảnh và thông tin văn bản thu được từ Web cho tra cứu ảnh trên Web.

Phản hồi liên quan là một quá trình trực tuyến mà cố gắng học mục đích của người dùng trong quá trình và là một công cụ mạnh được sử dụng truyền thống trong các hệ thống tra cứu thông tin [29]. Nó được giới thiệu đối với CBIR khoảng đầu những năm 1990, với mục đích mang người dùng vào lặp tra cứu để giảm khoảng cách ngữ nghĩa giữa những gì mà truy vấn biểu diễn và những gì người dùng nghĩ. Bằng việc tiếp tục học thông qua tương tác với các người dùng cuối, phản hồi liên quan đã được chỉ ra là cung cấp cải tiến hiệu năng đáng kể trong các hệ thống CBIR [30,31].

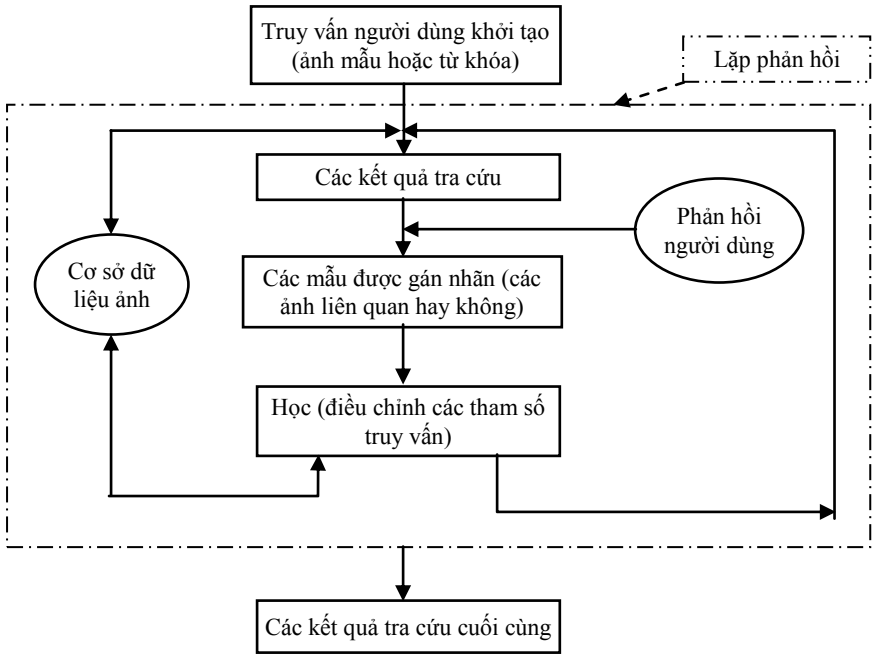
Một viễn cảnh tiêu biểu cho RF trong CBIR là như sau [33]:

(1) Hệ thống cung cấp các kết quả tra cứu khởi tạo thông qua truy vấn bởi mẫu, phác thảo,…

(2) Người dùng đánh giá các kết quả trên là có liên quan đến ảnh truy vấn hay không và độ liên quan là bao nhiêu.

(3) Thuật toán học máy được áp dụng để học phản hồi của người dùng. Sau đó quay về bước (2).

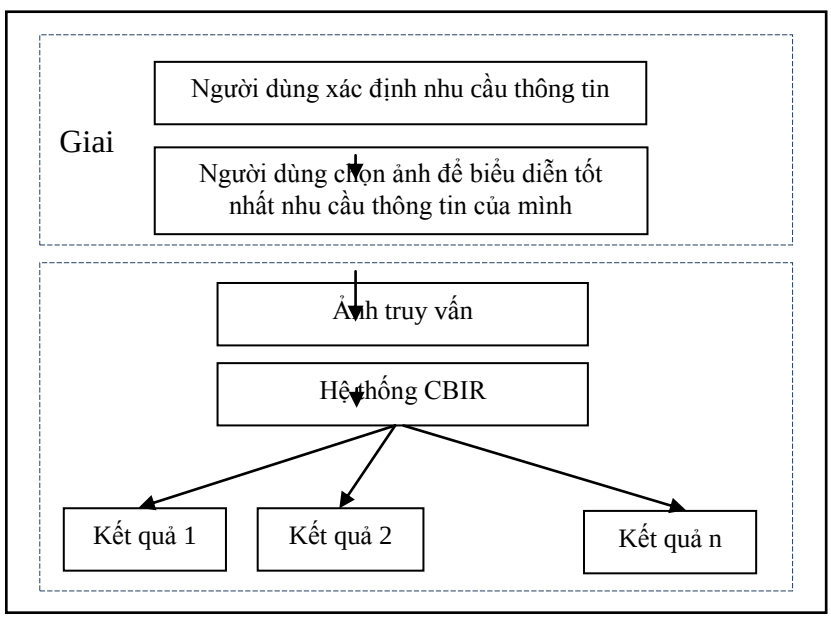
(2)-(3) được lặp cho đến khi người dùng thỏa mãn với các kết quả. Hình 1 chỉ ra một lược đồ đơn giản của một hệ thống CBIR với phản hồi liên quan.



Hình 1. Tra cứu ảnh dựa vào nội dung với phản hồi liên quan

Các đối tượng trả về so với truy vấn người dùng bởi nhiều hệ thống tra cứu ảnh dựa vào nội dung đã có thường không thỏa mãn nhu cầu thông tin của người dùng [7, 8, 9, 10]. Điều này là do một số lý do sau:

Lý do thứ nhất, nhu cầu thông tin của người dùng rất phong phú, vì thế khó có thể biểu diễn nhu cầu này với một ảnh truy vấn. Điều này sẽ rõ ràng hơn thông qua việc xét mô hình tra cứu tổng quát trong Hình 2. Để tra cứu theo mô hình tổng quát này, cần thực hiện hai giai đoạn như sau: Giai đoạn thứ nhất, người dùng xác định nhu cầu thông tin của mình (chẳng hạn nhu cầu muốn tìm tất cả những bông hoa hồng trong cơ sở dữ liệu), sau đó người dùng sẽ chọn ảnh truy vấn biểu diễn nhu cầu thông tin vừa xác định. Giai đoạn thứ hai, ảnh mà người dùng vừa chọn sẽ được sử dụng làm ảnh truy vấn và các phương pháp tra cứu ảnh khác nhau sẽ được thực hiện để cho ra tập các kết quả: *kết quả 1, kết quả 2,…. kết quả n*.



Hình 2. Mô hình tra cứu tổng quát

Chúng ta nhận thấy, trong mô hình tra cứu tổng quát trên Hình 2, nếu ảnh truy vấn không biểu diễn tốt nhu cầu thông tin rất phong phú của người dùng, cho dù các phương pháp tra cứu hiện nay có cho ra tập kết quả (gồm *kết quả 1, kết quả 2,…. kết quả n*) có độ chính xác 100% so với ảnh truy vấn (điều này không có trong thực tế) thì tập kết quả vẫn có thể không phải là mong muốn của người dùng. Chính vì lý do đó mà các phương pháp tra cứu sử dụng một ảnh truy vấn thường cho tập kết quả không đáp ứng kỳ vọng của người dùng; Lý do thứ hai, một ảnh thường gồm nhiều biểu diễn với độ quan trọng khác nhau nhưng các phương pháp thường coi độ quan trọng này là ngang nhau và lý do cuối cùng là các đặc trưng mức thấp không phản ánh được thông tin ngữ nghĩa của ảnh và hàm khoảng cách kết hợp với các đặc trưng mức thấp không thể hiện được nhận thức về độ tương tự trực quan của người dùng.

Các lý do ở trên là động lực để chúng tôi đề xuất phương pháp tra cứu ảnh có tên ERIN (Efficient Representation of Information Need) có ưu điểm biểu diễn tốt nhu cầu thông tin của người dùng do sử dụng nhiều ảnh và nhiều điểm để biểu diễn, xác định được độ quan trọng của mỗi biểu diễn ảnh và giảm khoảng cách ngữ nghĩa thông qua kỹ thuật phản hồi liên quan dẫn đến nâng cao chất lượng hệ thống tra cứu ảnh.

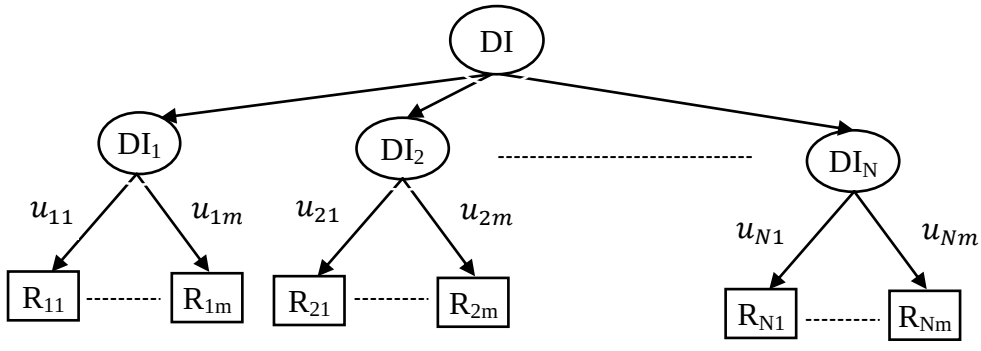
Phần còn lại của bài báo này được tổ chức như sau: trong phần 2, trình bày chi tiết phương pháp tra cứu ảnh ERIN dựa vào đa truy vấn và đa biểu diễn. Phần 3, trình bày thuật toán đề xuất cải tiến độ chính xác tra cứu sử dụng biểu diễn nhu cầu thông tin của người dùng hiệu quả. Phần 4, mô tả các kết quả thực nghiệm và cuối cùng là kết luận được đưa ra trong phần 5.

II. PHƯƠNG PHÁP TRA CỨU ẢNH DỰA VÀO ĐA TRUY VẤN VÀ ĐA BIỂU DIỄN

Từ một số ảnh do người dùng đưa vào làm truy vấn mà biểu diễn nhu cầu thông tin của họ, để có thể cho ra một tập các ảnh kết quả tương ứng với truy vấn đó, chúng ta cần có mô hình ảnh cơ sở dữ liệu, mô hình đa truy vấn và đa biểu diễn và mô hình tra cứu. Trong phần này, chúng tôi sẽ trình bày ba mô hình này, thuật toán xác định độ quan trọng biểu diễn và thuật toán tra cứu ảnh sử dụng đa truy vấn và đa biểu diễn.

Mô hình biểu diễn các ảnh trong cơ sở dữ liệu:

Trước khi tra cứu các ảnh, đầu tiên các ảnh trong tập ảnh phải được biểu diễn và lưu trữ trong cơ sở dữ liệu đặc trưng. Để thực hiện được việc đó, chúng ta cần có mô hình biểu diễn ảnh của tập ảnh. Kí hiệu mỗi ảnh trong cơ sở dữ liệu là DI_i, mỗi ảnh DI_i này sẽ có một tập các biểu diễn $\{R_{i1}^{DI_i}, R_{i2}^{DI_i}, \dots, R_{im}^{DI_i}\}$ với mỗi $R_{ij}^{DI_i}$ là một biểu diễn đặc trưng j của ảnh DI_i, mỗi biểu diễn này có một trọng số u_{ij} (được xác định qua thuật toán *IR* trên Hình 6) gắn với biểu diễn đặc trưng j của ảnh DI_i so với biểu diễn đặc trưng khác của ảnh DI_i. Hình 3 là một minh họa trực quan về mô hình này.



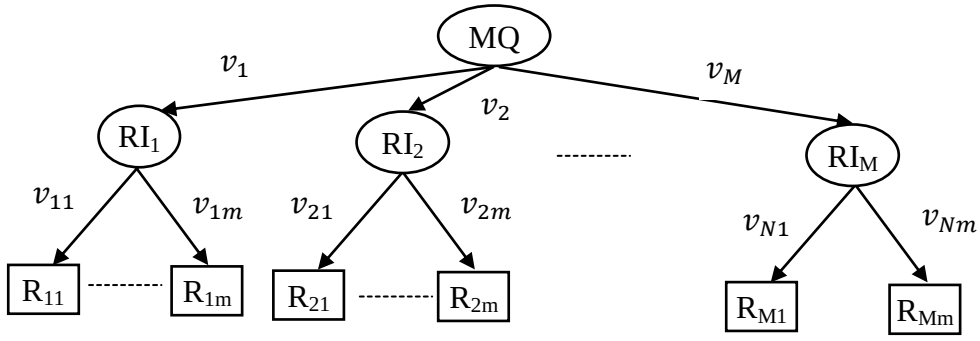
Hình 3. Mô hình biểu diễn ảnh cơ sở dữ liệu

Mô hình biểu diễn đa truy vấn và đa biểu diễn:

Sau khi đã có mô hình biểu diễn các ảnh cơ sở dữ liệu, bước tiếp theo, chúng ta cần có mô hình biểu diễn đa truy vấn và đa biểu diễn *MQ* (Multipoint Query). Mỗi truy vấn sẽ gồm *M* ảnh đại diện (được ký hiệu là RI_k), mỗi ảnh đại diện RI_k được biểu diễn tương tự như ảnh cơ sở dữ liệu, tức là mỗi ảnh RI_k này sẽ được biểu diễn bởi một tập các biểu diễn $\{R_{k1}^{RI_k}, R_{k2}^{RI_k}, \dots, R_{km}^{RI_k}\}$, mỗi $R_{kj}^{RI_k}$ là một biểu diễn đặc trưng j của ảnh RI_k, mỗi biểu diễn này có một trọng số v_{kj} gắn với biểu diễn đặc trưng j của ảnh RI_k so với biểu diễn đặc trưng khác của ảnh RIk.

Mô hình tra cứu đa truy vấn và đa biểu diễn:

Trong phương pháp này, một đại diện trong một truy vấn và một ảnh cơ sở dữ liệu có cùng cấu trúc. Độ tương tự giữa đa truy vấn và ảnh cơ sở dữ liệu được tính bằng tổng có trọng số của các độ tương tự biểu diễn đặc trưng riêng lẻ. Kết quả cuối cùng của tra cứu là một danh sách các ảnh được phân hạng theo thứ tự giảm dần của độ tương tự so với ảnh truy vấn. Cho *MQ* là một nút truy vấn và các RI_k với $k=1..M$ (các nút đại diện) là con của *MQ* . Cho $R_{k1}^{RI_k}, R_{k2}^{RI_k}, \dots, R_{km}^{RI_k}$ là con của RI_k (các nút biểu diễn đặc trưng). Cho v_i là trọng số của nút đại diện. Cho v_{kj} là các trọng số của các nút biểu diễn đặc trưng. Hình 4 là một minh họa trực quan về mô hình này.



Hình 4. Mô hình biểu diễn đa truy vấn và đa biểu diễn

Kí hiệu dis_{ik} là khoảng cách của một ảnh cơ sở dữ liệu thứ *i* đến một đại diện thứ *k* của truy vấn và được tính theo công thức (1) sau:

$$dis_{ik} = \sum_{j=1}^{m}(1 - u_{ij}v_{kj}) * distance(R_{ij}^{DI_i}, R_{kj}^{RI_k}) \quad (1)$$

Kí hiệu dis_i là khoảng cách của một ảnh cơ sở dữ liệu thứ *i* đến truy vấn và được tính theo công thức sau:

$$dis_i = min_{k=1..M}((1 - v_k) * dis_{ik}) \quad (2)$$

Trên cơ sở mô hình ảnh cơ sở dữ liệu, mô hình truy vấn đa điểm và mô hình tra cứu, chúng tôi xây dựng thuật toán tra cứu dựa vào đa truy vấn và đa biểu diễn. Thuật toán, có tên là **MQMRBR** (Multiple Queries and Multiple Representations Based Retrieval), tính khoảng cách giữa đa truy vấn và mỗi ảnh cơ sở dữ liệu, sau cho ra một danh sách được phân hạng theo thứ tự tăng dần của khoảng cách. Thuật toán **MQMRBR** được mô tả như trong Hình 5.

Thuật toán tra cứu ảnh dựa vào đa truy vấn và đa biểu diễn **MQMRBR** trên Hình 5 thực hiện như sau: Đầu tiên, pha xây dựng mô hình biểu diễn ảnh cơ sở dữ liệu được thực hiện. Trong pha này, mỗi ảnh DI_i trong tập ảnh cơ sở dữ liệu *DI* gồm *N* ảnh, thực hiện trích rút biểu diễn thứ *j* ($R_{ij}^{RI_i}$) của ảnh RI_i thông qua hàm ***Extracte***(). Đi cùng với biểu diễn thứ *j* này là một trọng số u_{ij} (để xác định độ quan trọng của biểu diễn thứ *j*, lúc ban đầu có độ quan trọng như nhau) cũng được gán thông qua thủ tục **Weight_Assign**(). Sau đó, pha xây dựng mô hình biểu diễn truy vấn đa điểm được tiến hành. Trong pha này, với mỗi ảnh RI_k trong tập *M* ảnh đại diện của truy vấn *MQ* do người dùng đưa vào sẽ có một trọng số v_k để xác định đại diện độ quan trọng của đại diện thứ *k*, trọng số này được tính toán thông qua thủ tục **RI_Weight_Compute**(). Trên mỗi ảnh RI_k, thực hiện trích rút biểu diễn thứ *j* ($R_{kj}^{RI_k}$) thông qua hàm ***Extracte***() và một trọng số tương ứng với biểu diễn này là v_{kj} (lúc ban đầu có độ quan trọng ngang nhau) cũng được tính toán thông qua hàm **R_Weight_Compute()**. Cuối cùng là pha tra cứu. Trong pha này, thực hiện tính khoảng cách giữa biểu diễn thứ *j* của ảnh DI_i ($R_{ij}^{DI_i}$) và ảnh RI_k ($R_{kj}^{RI_k}$) thông qua hàm ***distance*()** nhân với đối ngẫu của trọng số u_{ij} và v_{kj} để được khoảng cách giữa DI_i và RI_k, sau đó lưu vào dis_{ik}. Khoảng cách giữa một ảnh cơ sở dữ liệu và truy vấn đa điểm *MQ* là khoảng cách cực tiểu có trọng số của các khoảng cách riêng giữa ảnh cơ sở dữ liệu DI_i và từng ảnh đại diện RI_k của truy vấn, giá trị này được lưu trữ vào dis_i. Sau khi có khoảng cách của từng ảnh cơ sở dữ liệu DI_i với truy vấn đa điểm *MQ*, thủ tục **Sort()** sẽ sắp xếp các ảnh DI_i trong tập ảnh *DI* theo thứ tự tăng dần về khoảng cách so với truy vấn *MQ* và trả về tập ảnh kết quả *S*.

Thuật toán MQMRBR *(*Multiple Queries and Multiple Representations Based Retrieval)

Input:
Tập *N* ảnh cơ sở dữ liệu *DI*
Tập *M* ảnh truy vấn *MQ*
Số đặc trưng *m*
Ouput:
Tập ảnh kết quả *S*

1. Xây dựng mô hình biểu diễn ảnh cơ sở dữ liệu
For *i*←*1* **to** *N* **do**
For *j*←*1* **to** *m* **do**
{
$R_{ij}^{DI_i}$←***Extracte***(DI_i) // thực hiện trích rút biểu diễn đặc trưng *j* của ảnh cơ sở dữ liệu thứ *i*
Weight_Assign(u_{ij}) // lúc đầu gán trọng số 1 cho các biểu diễn đặc trưng thứ *j* của ảnh cơ sở dữ liệu thứ *i*
}

2. Xây dựng mô hình biểu diễn truy vấn đa điểm
For *k=1* **to** *M* **do**
For *j=1* **to** *m* **do**
{
$R_{kj}^{RI_k} \leftarrow \boldsymbol{Extracte}(RI_k)$ // trích rút biểu diễn đặc trưng j của ảnh đại diện thứ k thuộc truy vấn đa điểm
RI_Weight_Compute(v_k) // tính trọng số cho ảnh đại diện thứ k của truy vấn đa điểm
R_Weight_Compute(v_{kj}) // lúc đầu gán trọng số 1 cho các biểu diễn đặc trưng thứ j của ảnh đại diện thứ k
}
3. Thực hiện mô hình tra cứu truy vấn đa điểm
For *i←1* **to** *N* **do**
{
$dis_i \leftarrow 0$
For *k←1* **to** *M* **do**
{
$dis_{ik} \leftarrow 0$
For *j←1* **to** *m* **do**
$dis_{ik} \leftarrow dis_{ik} + (1 - u_{ij}v_{kj}) * \boldsymbol{distance}(R_{ij}^{DI_i}, R_{kj}^{RI_k})$
}
$dis_i \leftarrow min_{k=1..M}((1 - v_{k)} * dis_{ik)}$
}
Sort(*DI*) // sắp xếp các ảnh trong tập ảnh *DI* theo thứ tự tăng dần của khoảng cách so với truy vấn đa điểm *MQ*.
Return *S* // danh sách các ảnh có khoảng cách nhỏ nhất so với *MQ*

Hình 5. Thuật toán tra cứu ảnh dựa vào đa truy vấn và đa biểu diễn **MQMRBR**

III. CẢI TIẾN ĐỘ CHÍNH XÁC TRA CỨU

Trong số *k* ảnh được trả về bởi việc thực hiện đa truy vấn và đa biểu diễn trong thuật toán **MQMRBR**, người dùng sẽ chọn *n* ảnh liên quan. Dựa vào n điểm liên quan này, chúng ta gọi thuật toán IR để xác định độ quan trọng biểu diễn. Một số các đại diện sẽ được tính toán trong số n ảnh liên quan để xây dựng đa truy vấn. Việc tính toán các đại diện được thực hiện bằng cách phân cụm tập *n* đối tượng ảnh liên quan và chọn trọng tâm của các cụm làm các đại diện. Thuật toán phân cụm được sử dụng là thuật toán trong [11], có tính chất bảo toàn được trọng tâm và do đó đảm bảo rằng các đại diện được lựa chọn là các điểm từ tập liên quan. Thuật toán nhận đầu vào là *M* cụm mong muốn, khi các điểm mới được thêm vào, thuật toán tính toán một tập các cụm và duy trì số cụm nhỏ hơn hoặc bằng *M*. Tiếp theo, phương pháp tính khoảng cách giữa từng đối tượng ảnh và đa truy vấn để cho ra một danh sách được phân hạng theo thứ tự tăng dần của khoảng cách so với truy vấn. Quá trình trên được lặp lại cho đến khi người dùng dừng phản hồi. Trọng số v_k tương ứng với mỗi đại diện RI_k của đa truy vấn là số các đối tượng ảnh liên quan trong cụm tương ứng.

Thuật toán tính độ quan trọng của biểu diễn:

Mỗi một ảnh gồm nhiều biểu diễn được biểu diễn bởi một điểm trong không gian đặc trưng. Thông thường, các phương pháp coi các biểu diễn này có độ quan trọng như nhau. Điều này không phản ảnh đúng thực tế là có một số biểu diễn quan trọng hơn các biểu diễn còn lại. Do đó, chúng tôi quan tâm tới việc xác định độ quan trọng của mỗi biểu diễn của ảnh.

Ý tưởng chính của việc xác định độ quan trọng biểu diễn là dựa vào sự phản hồi của người dùng. Khi người dùng phản hồi một số ảnh là liên quan ngữ nghĩa với ảnh truy vấn, chúng tôi sẽ coi mỗi ảnh là một điểm dữ liệu trong không gian và xét hình bao các điểm dữ liệu này. Một hình bao các điểm như thế sẽ được chiếu xuống các trục tương ứng với các biểu diễn, sau đó tính phương sai của các điểm này theo mỗi trục (sẽ biết được độ phân tán dữ liệu theo một trục trong không gian lớn cũng có nghĩa là độ quan trọng theo trục đó nhỏ). Do đó, độ quan trọng của mỗi biểu diễn trong không gian là nghịch đảo của phương sai của các điểm theo trục đó.

Chúng tôi sẽ trình bày thuật toán IR (Importance of Representation) để xác định độ quan trọng biểu diễn. Thuật toán tính độ quan trọng của biểu diễn trong không gian biểu diễn *RS*. Hình 6 dưới đây là thuật toán IR.

Thuật toán *IR* - Importance of Representation

Input:
Tập n điểm dữ liệu *C*
Tập các biểu diễn *RS*
Số biểu diễn *m*
Ouput:
Trọng số của biểu diễn thứ j $Weight_j$

For $j \leftarrow 1$ to m do
{
$\mu_j \leftarrow \frac{1}{n}\sum_{i=1}^{n} C_i^{RS}[j]$
$\sigma_j^F \leftarrow (\frac{1}{n}\sum_{i=1}^{n}(C_i^{RS}[j] - \mu_j)^2)^{1/2}$
$Weight_j \leftarrow \frac{1}{\sigma_j^{RS}}$ // trọng số của biểu diễn thứ j
}

Hình 6. Thuật toán tính độ quan trọng của biểu diễn IR

Thuật toán IR trên Hình 6, lấy đầu vào là *n* điểm (ảnh) $C_1^{RS}, \dots C_i^{RS}, \dots C_n^{RS}$ trong một cụm trên không gian *RS*. Lúc này, theo đặc trưng thứ *j* của không gian *RS* sẽ có *n* điểm dữ liệu $C_1^{RS}[j], \dots C_i^{RS}[j], \dots C_n^{RS}[j]$ và thuật toán tính phương sai σ_j^{RS} của n điểm dữ liệu này theo trục *j* của không gian *RS*. Sau khi tính được giá trị của phương sai σ_j^{RS}, thuật toán đưa ra độ quan trọng của từng biểu diễn *j* trong không gian *RS*. Độ quan trọng của biểu diễn theo trục *j* sẽ được tính bằng $\frac{1}{\sigma_j^{RS}}$ và gán cho $Weight_j$.

Hình 7 dưới đây là mô tả thuật toán tra cứu ảnh sử dụng biểu diễn nhu cầu thông tin người dùng hiệu quả có tên ERIN (Efficient Representation of Information Need).

Thuật toán tra cứu ảnh sử dụng biểu diễn nhu cầu thông tin người dùng hiệu quả, có tên ERIN trên Hình 7, được thực hiện như sau: Khi người dùng gửi một tập ảnh làm đa truy vấn *MQ*, phương pháp sẽ sử dụng thuật toán ***MQMRBR*** để tra cứu trên tập các ảnh cơ sở dữ liệu *DI* và cho kết quả là tập các ảnh *S*. Người dùng thực hiện việc chọn tập các ảnh liên quan *E* trong tập *S* thông qua hàm ***User_Choose_RelevanceImage***(), phương pháp sẽ phân cụm tập *E* này thành *M* cụm thông qua hàm ***Clustering***() và gán cho *C*, tâm của m cụm được tính toán thông qua hàm ***Compute_Centroid***() và gán cho tập đại diện *RI*. Trọng số cho tâm cụm thứ *k* được tính thông qua hàm ***RI_Weight_Compute***() và trọng số v_{kj} cho các biểu diễn thứ *j* của tâm cụm thứ *k* được tính qua hàm ***IR***(). Khoảng cách giữa ảnh cơ sở dữ liệu DI_i và truy vấn *MQ* được tính theo công thức (1) và (2). Quá trình này tiếp tục cho đến khi người dùng dừng việc chọn các ảnh liên quan.

Thuật toán ERIN (Efficient Representation of Information Need)

Input:
Tập N ảnh cơ sở dữ liệu — DI
Tập M ảnh đại diện truy vấn — MQ
Tập biểu diễn — RS
Số biểu diễn — m

Ouput:
Tập ảnh kết quả — S'

MQMRBR(DI, MQ, S) // Thực hiện trên tập ảnh DI với truy vấn đa điểm MQ để cho ra tập kết quả S
Repeat
$E \leftarrow$ ***User_Choose_RelevanceImage***(S, n) // người dùng chọn các ảnh liên quan từ tập ảnh S
$C \leftarrow$ ***Clustering***(E, M) // phân tập ảnh liên quan E thành M cụm
RI $\leftarrow$ ***Compute_Centroid***(C, M)
For $k \leftarrow 1$ **to** M **do**
For $j \leftarrow 1$ **to** m **do**
{
RI_Weight_Compute(v_k) // tính trọng số cho tâm cụm thứ k của truy vấn đa điểm
IR(E, RS, m, $Weight_j$) tính trọng số cho các biểu diễn đặc trưng thứ j của tâm cụm thứ k
$v_{kj} \leftarrow Weight_j$
}
Tính dis_{ik} theo công thức (1):

$$dis_{ik} \leftarrow \sum_{j=1}^{m} (1 - u_{ij} v_{kj)} * distance(R_{ij}^{DI_i}, R_{kj}^{RI_k})$$

Tính dis_i theo công thức sau (2):

$$dis_i = min_{k=1..M}((1 - v_k) * dis_{ik})$$

Sort(DI) // sắp xếp các ảnh trong tập ảnh cơ sở dữ liệu DI theo thứ tự tăng dần của khoảng cách so với truy vấn đa điểm MQ.
Return S' // k ảnh có khoảng cách nhỏ nhất với MQ
Until (User dừng phản hồi)

Hình 7. Thuật toán tra cứu ảnh sử dụng biểu diễn nhu cầu thông tin người dùng hiệu quả ERIN

IV. THỰC NGHIỆM

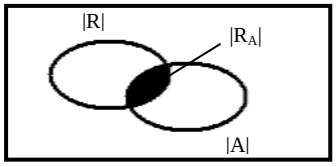
Hệ thống được cài đặt trên máy tính PC Pentium G3220 3.00 GHz chạy hệ điều hành Windows 8.1 với một cơ sở dữ liệu ảnh gồm 10.800 ảnh[1]. Các ảnh được lưu trữ theo định dạng JPEG với cỡ 120×80 và được lượng hóa thành 16 màu. Cơ sở dữ liệu bao gồm 80 chủ đề: biển, thẻ, ngựa, bướm, hoa, vận động viên thể thao, lướt ván, thuyền buồm, hoa quả, cờ, chim, nhà, thác nước, gấu, linh dương đầu bò, ô tô, núi - hoàng hôn, cánh rừng,…. Cơ sở dữ liệu này sẽ được sử dụng để minh chứng sự chính xác của kỹ thuật. Các ảnh trong 50 ảnh được tra cứu đầu tiên được phân thành các mẫu tích cực và tiêu cực (theo đánh giá của người dùng).

Chúng tôi đã so sánh nghiên cứu của chúng tôi với nghiên cứu CBsIR [13] và CCH [12]. Để cung cấp các kết quả đáng tin cậy, 5 ảnh từ mỗi trong mười chín loại ở trên được chọn ra ngẫu nhiên làm các ảnh truy vấn. Đồ thị triệu hồi chính xác [14] được sử dụng để so sánh ERIN, CBsIR và CCH. Sự chính xác là chính xác trung bình của tổng các truy vấn và kết quả được chỉ ra trong Hình 9.

Ký hiệu R là một tập các ảnh liên quan trong cơ sở dữ liệu, A là tập các ảnh tra cứu được trả về, R_A là tập các ảnh liên quan trong tập A (Hình 8).

Triệu hồi (Recall) là tỷ số của các ảnh liên quan trong cơ sở dữ liệu được tra cứu theo một truy vấn. Độ chính xác (Precision) là tỷ số của các ảnh được tra cứu mà liên quan đến ảnh truy vấn.

$$recall = \frac{area(R_A)}{area(R)} \ , \quad precision = \frac{area(R_A)}{area(A)}$$



Hình 8. Triệu hồi và chính xác cho các kếquả truy vấn.

Ảnh hoàng hôn được sử dụng như ảnh truy vấn của ERIN, CBsIR và CCH để chỉ ra hiệu quả của ERIN.

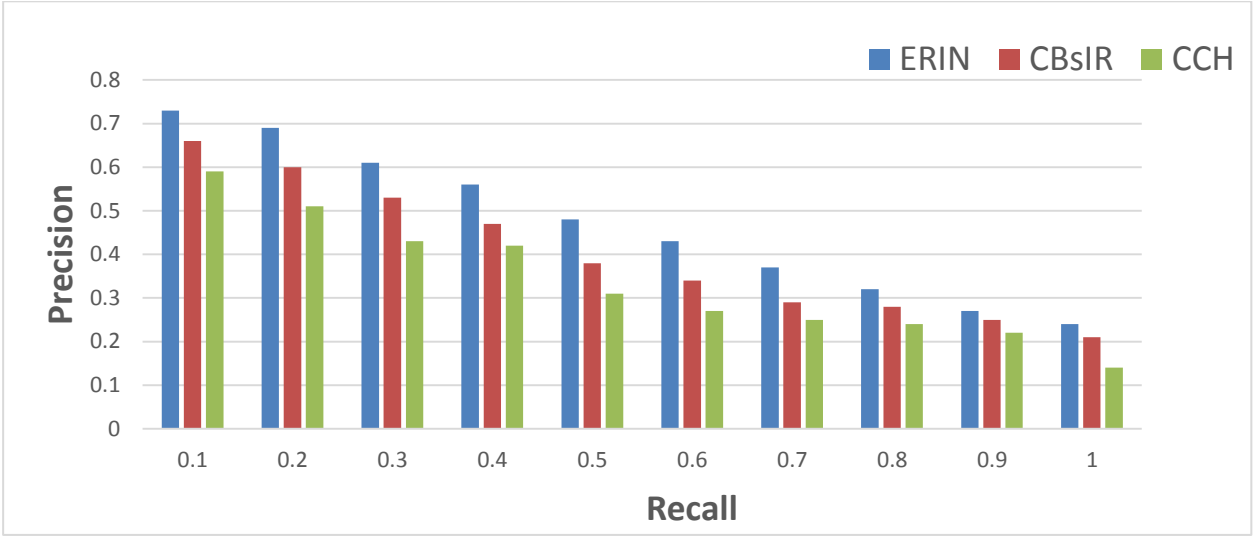
Bảng 1 đưa ra tóm tắt các kết quả trung bình truy vấn. Các kết quả tra cứu được tóm tắt dưới dạng triệu hồi chính xác. Trong truy vấn hai thực nghiệm được thực hiện, thứ nhất kỹ thuật ERIN được sử dụng cho quá trình tra cứu. Kỹ thuật CBsIR được sử dụng trong thực nghiệm thứ hai và cuối cùng là kỹ thuật CCH.

Bảng 1. Các kết quả trung bình của truy vấn.

Recall	***Precision***		
	ERIN	***CBsIR***	***CCH***
0.1	0.73	0.66	0.59
0.2	0.69	0.6	0.51
0.3	0.61	0.53	0.43
0.4	0.56	0.47	0.42
0.5	0.48	0.38	0.31
0.6	0.43	0.34	0.27
0.7	0.37	0.29	0.25
0.8	0.32	0.28	0.24
0.9	0.27	0.25	0.22
1	0.24	0.21	0.14

Hình 9 chỉ ra kết quả của ERIN tốt hơn ***CBsIR*** và CCH.

[1] https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval



Hình 9. So sánh Precision - Recall của **ERIN** với **CBsIR** và **CCH**.

V. KẾT LUẬN

Chúng tôi đã tập trung vào đề xuất phương pháp, có tên là ERIN, giải quyết ba vấn đề chính đó là: (1) biểu diễn tốt nhu cầu thông tin của người dùng, (2) xác định được độ quan trọng của mỗi biểu diễn và giảm khoảng cách ngữ nghĩa giữa đặc trưng mức thấp và khái niệm mức cao. Để giải quyết được vấn đề (1), chúng tôi đã sử dụng nhiều truy vấn để biểu diễn thông tin của người dùng. Với vấn đề (2) chúng tôi đã tận dụng sự đánh giá của người dùng để xác định độ quan trọng của mỗi biểu diễn đặc trưng và với vấn đề (3) chúng tôi sử dụng kỹ thuật phản hồi liên quan của người dùng để giải quyết. Các kết quả thực nghiệm trên cơ sở dữ liệu gồm 10.800 ảnh chỉ ra độ chính xác của phương pháp được đề xuất. Thực nghiệm cũng chỉ ra hiệu năng của ERIN cao hơn phương pháp CBsIR và CCH.

TÀI LIỆU THAM KHẢO

[1] J. Eakins, M. Graham, Content-based image retrieval, Technical Report, University of Northumbria at Newcastle, 1999.

[2] A. Mojsilovic, B. Rogowitz, Capturing image semantics with low-level descriptors, Proceedings of the ICIP, September 2001,pp. 18–21.

[3] X.S. Zhou, T.S. Huang, CBIR: from low-level features to high-level semantics, Proceedings of the SPIE, Image and Video Communication and Processing, San Jose, CA, vol. 3974, 2000, pp. 426–431.

[4] A. Pentland, R. W. Picard, and S. Sclaroff (1996). Photobook: content-based manipulation for image databases.International Journal of Computer Vision, 18(3):233–254.

[5] Y. Chen, J.Z. Wang, R.Krovetz, An unsupervised learning approach to content-based image retrieval, IEEE Proceedings of the International Symposium on Signal Processing and its Applications, July 2003, pp. 197–200.

[6] C. Carson, S. Belongie, H. Greenspan, and J. Malik (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8):1026–1038

[7] K. Chakrabarti, K. Porkaew, and S. Mehrotra (2000). Efficient query refinement in multimedia databases. Proceedings of International Conference in Data Engineering(ICDE).

[8] Y. Ishikawa, R. Subramanya, and C. Faloutsos (1998). Mindreader: Querying databases through multiple examples. Proc. Of VLDB.

[9] K. Porkaew, K. Chakrabarti, and S. Mehrotra (1999). Query refinement for content-based multimedia retrieval in MARS. Proceedings of ACM Multimedia Conference.

[10] Y. Rui, T. Huang, and S. Mehrotra (1998). Relevance feedback techniques in interactive content-based image retrieval. Proc. of IS&T and SPIE Storage and Retrieval of Image and Video Databases.

[11] M. Charikar, C. Chekuri, T. Feder, and R. Mot-wani (1997). Incremental clustering and dynamic infor-mation retrieval. Proc. of ACM Symposium on Theory of Computing.

[12] R.O Stehling, M.A. Nascimento, A.X. Falc˜ao (2003), "Cell histograms versus color histograms for image representation and retrieval", *Knowledge and Information Systems (KAIS) Journal*, pp. 151-179.

[13] Luo, Jie and Nascimento, Mario A. (2004), Content Based Sub Image Retrieval Using Relevance Feedback,Proceedings of the 2Nd ACM International Workshop on Multimedia Databases.

[14] B. Yates and R. Neto (1999), Modern Information Retrieval, Addison Wesley.

[15] 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.

[16] A. Gupta and R. Jain (1997). Visual information retrieval. Communications of the ACM, 40(5):70–79.

[17] L. Chen, M. T.¨Ozsu, and V. Oria (2004). MINDEX: An efficient index structure for salient-object-based queries in video databases. Multimedia Systems, 10 (1):56–71.

[18] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N.Yianilos (2000). The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. IEEE Transactions on Image Processing, 9(1):20–37.

[19] [Flickner et al., 1995] 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.

[20] K. A. Hua, N. Yu, and D. Liu (2006). Query Decomposition: A Multiple Neighborhood Approach to Relevance Feedback Processing in Content-based Image Retrieval. InProceedings of the IEEE ICDE Conference.

[21] K. Vu, K. A. Hua, and W. Tavanapong (2003). Image retrieval based on regions of interest. IEEE Transactions on Knowledge and Data Engineering, 15(4):1045–1049.

[22] J. Z. Wang, J. Li, and G. Wiederhold,(2001). "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 23, no. 9, pp. 947-963.

[23] W. Y. Ma and B. Manjunath (1997). Netra: a toolbox for navigating large image databases. In Proceedings of the IEEE International conference on Image Processing, pages 568–571

[24] V. Ogle and M. Stonebraker (1995). Chabot: retrieval from a relational database of images.IEEE Computer, 28(9):40–48.

[25] M. Ortega-Binderberger and S. Mehrotra (2004). Relevance feedback techniques in the MARS image retrieval systems. Multimedia Systems, 9(6):535–547.

[26] H. T. Shen, B. C. Ooi, and X. Zhou (2005). Towards effective indexing for very large video sequence database. In Proceedings of the ACM SIGMOD Conference, pages 730–741.

[27] 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.

[28] Smith, J.R., Chang, S.F., (1996). VisualSEEk: A fully automated content-based image query system.

In: Proceedings of the ACM Int'l Multimedia Conference, pp. 87–98.

[29] G. Salton, Automatic Text Processing, Addison-Wesley, Reading, MA, 1989.

[30] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback:a power tool for interactive content-based image retrieval, IEEE Trans. Circuits Video Technol. 8 (5) (1998) 644–655.

[31] Y. Rui, T.S. Huang, Optimizing learning in image retrieval, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, June 2000, pp. 1236–1243.

[32] Brunelli, R., Mich, O., (2000). Image retrieval by examples. IEEE Transactions on Multimedia 2 (3), 164–171.

[33] X.S. Zhu, T.S. Huang, Relevance feedback in image retrieval: a comprehensive review, Multimedia System 8 (6) (2003) 536–544.

[34] Y. Rui, T.S. Huang, S.-F. Chang, Image retrieval: current techniques, promising directions, and open issues, J. Visual Commun. Image Representation 10 (4) (1999) 39–62.

[35] Y. Yan, M.-L. Shyu, and Q. Zhu (2016), Negative correlation discovery for big multimediadata semantic concept mining and retrieval, in Proceedings of the IEEE international Conference on Semantic Computing, pp. 55-62.

AN IMAGE RETRIEVAL METHOD EFFICIENTLY REPRESENTS THE USER'S INFORMATION NEED

Nguyen Huu Quynh, Dao Thi Thuy Quynh, Ngo Quoc Tao, Cu Viet Dung, Phương Văn Cảnh, An Hong Son

ABSTRACT *— Most of the conventional approaches to content-based on image retrieval is not efficiently represents the user's information need. The reasons for these limitations are: (a) the user's information needs are very rich, so it is difficult to perform this with a query image, (b) an image usually includes multiple representations with different importance but these methods are often considered equal importance, (c) low-level image features do not capture the semantics of images , (d) distance function associated with low level features can not express user's perception of visual similarity. . In order to overcome these problems, we propose image retrieval method, called ERIN (Efficient Representation of Information Need). The method has the advantages that efficently represent the user's information need by using multiple images and Multiple Representations. Beside, the method determines the importance of each representation which leads to improve quality content – based on image retrieval system. Our experimental results on a database of over 10.800 images. The experimental results indicate that this technique improved the performance of content – based on image retrieval compared to established methods and the results are closed to the user's information need.*

Keywords— *Content based image retrieval, representation of information need, multiple queries, and multiple representations, feature vector.*