

MỘT PHƯƠNG PHÁP TRA CỨU ẢNH KHAI THÁC HIỆU QUẢ THÔNG TIN PHẦN HỒI TỪ NGƯỜI DÙNG

Cù Việt Dũng¹, Nguyễn Hữu Quỳnh¹, Ngô Quốc Tạo², An Hồng Sơn³, Lê Mạnh Hùng⁴

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

²Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam,

³Trường Đại học Công nghiệp Việt Hưng,

⁴Phòng Khảo thí và kiểm định chất lượng, Trường Đại học Điện lực,

quynhnh@epu.edu.vn, dungcv@epu.edu.vn, nqtạo@ioit.ac.vn, sonanhongvnh@gmail.com, hunglm@gmail.com

TÓM TẮT: Đã có nhiều phương pháp tra cứu ảnh phân hồi liên quan sử dụng máy vectơ hỗ trợ (SVM). Tuy nhiên, các phương pháp này chưa quan tâm đến sự cân bằng của hai lớp trong quá trình phân hồi của người dùng dẫn đến độ chính xác thấp. Trong bài báo này, chúng tôi đề xuất một phương pháp tra cứu ảnh phân hồi liên quan hiệu quả, gọi là RFBC (Image retrieval using relevance feedback with balanced classes), cho phép nâng cao độ chính xác hệ thống tra cứu ảnh thông qua việc cân bằng số mẫu trong mỗi lớp của tập ví dụ huấn luyện. Phương pháp của chúng tôi tận dụng sự đánh giá của người dùng để xác định vị trí có thể thu được hai lớp có số lượng mẫu cân bằng. Chúng tôi cũng cung cấp các kết quả thực nghiệm trên cơ sở dữ liệu đặc trưng gồm 10800 ảnh để chỉ ra độ chính xác của phương pháp.

Từ khóa: Tra cứu ảnh dựa vào nội dung, không gian đặc trưng, truy vấn tối ưu, phân hồi liên quan với các lớp cân bằng.

I. GIỚI THIỆU

Tra cứu ảnh dựa vào nội dung (CBIR) đã thu hút nhiều sự quan tâm trong những thập kỷ qua [1, 2, 3, 4, 5]. Các hệ thống CBIR truyền thống thường sử dụng độ đo khoảng cách Euclidean trong một không gian đặc trưng mức thấp nhiều chiều để đo độ tương tự giữa ảnh truy vấn và các ảnh trong cơ sở dữ liệu [1, 2, 3, 4, 6]. Tuy nhiên, độ đo khoảng cách Euclidean trong không gian nhiều chiều thường không hiệu quả do khoảng trống giữa các đặc trưng mức thấp và các khái niệm ngữ nghĩa mức cao.

Để thu hẹp khoảng trống ngữ nghĩa, phân hồi liên quan được giới thiệu như một công cụ mạnh để tăng cường hiệu năng của CBIR [7,8]. Huang và cộng sự đã giới thiệu cả hai kỹ thuật đánh lại trọng số và dịch chuyển điểm truy vấn [9,10]. Một ánh xạ tự tổ chức được sử dụng để xây dựng các thuật toán RF [11]. Trong [12], máy vectơ hỗ trợ (SVM) một lớp đánh giá mật độ của các mẫu phân hồi tích cực. Bắt nguồn từ SVM một lớp, một biased SVM (BSVM) thừa hưởng sự thành công của SVM một lớp nhưng kết hợp các mẫu phân hồi tiêu cực [13]. Xem xét cấu trúc hình học của các đặc trưng trực quan mức thấp [14] và [15] đã đề xuất các cách tiếp cận dựa vào học đa dạng để tìm cấu trúc nội tại của các ảnh và cải tiến hiệu năng tra cứu. Với quan sát rằng “tất cả các mẫu tích cực là như nhau; mỗi mẫu tiêu cực là tiêu cực theo cách riêng của nó”, RF được phát biểu như bài toán học không gian con lệch (biased), trong đó có một số các lớp chưa biết, nhưng người dùng chỉ đề cập đến lớp tích cực [16, 17, 18]. Tuy nhiên, tất cả các phương pháp này có một số giới hạn. Chẳng hạn, phương pháp trong [9] và [10] được dựa vào kinh nghiệm, phương pháp ước lượng mật độ trong [12] bỏ qua mọi thông tin chứa trong các mẫu phân hồi tiêu cực và các kỹ thuật học không gian con phân biệt trong [16] và [18] thường bị vấn đề “cỡ mẫu nhỏ”. Xem các phân hồi tích cực và tiêu cực như hai nhóm khác nhau, các RF dựa vào phân lớp [19, 20, 21] đã trở thành một kỹ thuật phổ biến trong cộng đồng CBIR. Tuy nhiên, RF rất khác so với bài toán phân lớp truyền thống bởi vì các phân hồi được cung cấp bởi người dùng thường bị giới hạn trong các hệ thống tra cứu ảnh thực. Do đó, các phương pháp học mẫu nhỏ là hứa hẹn cho RF.

SVM hai lớp là một trong các phương pháp học mẫu nhỏ phổ biến được sử dụng rộng rãi trong những năm gần đây và thu được hiệu năng tốt trong phân lớp vì khả năng tổng quát của nó [20, 21, 22, 23, 24]. SVM có thể thu được một rủi ro cấu trúc cực tiểu bằng việc cực tiểu các chiều Vapnik-Chervonenkis [23]. Guo và cộng sự đã phát triển một độ đo tương tự có ràng buộc cho tra cứu ảnh [22], học một biên mà chia các ảnh thành hai nhóm và các mẫu bên trong biên được phân hạng bởi khoảng cách Euclidean đối với ảnh truy vấn. Phương pháp học SVM tích cực lựa chọn các mẫu gần với biên như các mẫu thông tin nhất cho người dùng gắn nhãn [24]. Các kỹ thuật lấy mẫu ngẫu nhiên được áp dụng để giảm sự không ổn định, độ lệch và các vấn đề quá khớp trong phân hồi liên quan với SVM [21]. Li và cộng sự đã đề xuất một phương pháp SVM đa huấn luyện bằng việc thích nghi một kỹ thuật đồng huấn luyện và một phương pháp lấy mẫu ngẫu nhiên [25]. Tuy nhiên, hầu hết các cách tiếp cận SVM bỏ qua sự khác nhau cơ bản giữa hai nhóm phân hồi riêng biệt, tức là tất cả các phân hồi tích cực cùng một khái niệm tương tự trong khi mỗi phân hồi tiêu cực thường thuộc về các khái niệm khác nhau. Các kỹ thuật phân hồi liên quan SVM truyền thống coi các phân hồi tích cực và tiêu cực là ngang bằng nhau [20, 21, 22, 24, 25]. Sử dụng trực tiếp SVM như một lược đồ RF sẽ gây nguy hại tiềm năng đến hiệu năng của các hệ thống CBIR. Một vấn đề bắt nguồn từ thực tế rằng các khái niệm ngữ nghĩa khác nhau ở trong các không gian con khác nhau và nó là mục đích của các lược đồ RF để tìm ra “không gian con nào” [16]. Tuy nhiên, nó sẽ là một gánh nặng cho các lược đồ RF dựa vào SVM để điều chỉnh các tham số cục bộ để thích nghi các thay đổi của không gian con. Các khó khăn đó làm giảm độ chính xác của các cách tiếp cận SVM truyền thống cho CBIR. Ngoài ra, nó là vấn đề để kết hợp thông tin của các mẫu chưa được gắn nhãn vào các lược đồ RF dựa vào SVM

truyền thống cho CBIR, mặc dù các mẫu chưa gán nhãn là rất hữu ích trong xây dựng bộ phân lớp tối ưu, giảm nhiễu và tăng cường hiệu năng của hệ thống.

Ngoài các giới hạn như đã được trình bày ở trên, các nghiên cứu ở trên chưa quan tâm đến vấn đề cân bằng số các mẫu của hai lớp trong tập ví dụ huấn luyện. Hạn chế này là một trong những nguyên nhân chính dẫn đến hiệu năng của hệ thống chưa cao. Chúng tôi sẽ chỉ ra hạn chế này thông qua thực nghiệm như sau:

Chúng tôi tiến hành thực nghiệm với 30 truy vấn (được lấy ngẫu nhiên trong cơ sở dữ liệu đặc trưng Corel¹). Mỗi truy vấn được thực hiện hai lần lặp phân hồi sử dụng SVM [26]. Với mỗi truy vấn, chúng tôi thu được hai lớp và số các mẫu tương ứng với mỗi lớp cùng với tỉ lệ chênh lệch giữa hai lớp (là tỉ số giữa số lượng chênh lệch giữa hai lớp với số lượng của lớp lớn nhất) như được chỉ ra trong Bảng 1.

Bảng 1. Tỉ lệ chênh lệch giữa hai lớp trong tập ví dụ huấn luyện thu được tương ứng với mỗi truy vấn.

STT	ID ảnh truy vấn	Số mẫu lớp tích cực	Số mẫu lớp tiêu cực	Tỉ lệ chênh lệch (%)	STT	ID ảnh truy vấn	Số mẫu lớp tích cực	Số mẫu lớp tiêu cực	Tỉ lệ chênh lệch (%)
1	8114	17	83	79.52	16	3775	47	53	11.32
2	2755	27	73	63.01	17	2121	4	96	95.83
3	5464	5	95	94.74	18	2708	16	84	80.95
4	7548	46	54	14.81	19	6643	3	97	96.91
5	9619	4	96	95.83	20	5103	46	54	14.81
6	10356	42	58	27.59	21	3791	13	87	85.06
7	5907	17	83	79.52	22	8956	5	95	94.74
8	1497	8	92	91.30	23	6308	14	86	83.72
9	1612	78	22	71.79	24	5925	59	41	30.51
10	2779	14	86	83.72	25	9884	9	91	90.11
11	9072	1	99	98.99	26	3080	28	72	61.11
12	2744	23	77	70.13	27	8159	3	97	96.91
13	8785	15	85	82.35	28	8120	23	77	70.13
14	2627	25	75	66.67	29	4099	31	69	55.07
15	10024	3	97	96.91	30	6117	8	92	91.30

Từ Bảng 1, chúng ta thấy rằng, các truy vấn có tỉ lệ chênh lệch của hai lớp cao là 5464, 9619, 1497, 9072, 10024, 2121, 6643, 8956, 9884, 8159, 6117. Các truy vấn có tỉ lệ chênh lệch của hai lớp thấp là 7548, 10356, 3775, 5103, 5925. Thực hiện tra cứu ảnh sử dụng bộ phân lớp SVM đối với mỗi truy vấn, chúng ta nhận được các kết quả tra cứu như trong Bảng 2 ở dưới:

Bảng 2. Độ chính xác² tra cứu của 30 truy vấn trong Bảng 1

STT	ID ảnh truy vấn	Tỉ lệ chênh lệch (%)	Độ chính xác (%)	STT	ID ảnh truy vấn	Tỉ lệ chênh lệch (%)	Độ chính xác (%)
1	8114	79.52	28	16	3775	11.32	50
2	2755	63.01	39	17	2121	95.83	4
3	5464	94.74	6	18	2708	80.95	22
4	7548	14.81	68	19	6643	96.91	8
5	9619	95.83	5	20	5103	14.81	85
6	10356	27.59	77	21	3791	85.06	14
7	5907	79.52	27	22	8956	94.74	7
8	1497	91.30	9	23	6308	83.72	28
9	1612	71.79	92	24	5925	30.51	86
10	2779	83.72	19	25	9884	90.11	11
11	9072	98.99	1	26	3080	61.11	47
12	2744	70.13	42	27	8159	96.91	4
13	8785	82.35	19	28	8120	70.13	36
14	2627	66.67	48	29	4099	55.07	43
15	10024	96.91	5	30	6117	91.30	17

Chúng ta thấy rằng, trong Bảng 2, các truy vấn có tỉ lệ chênh lệch giữa hai lớp cao như 5464, 9619, 1497, 9072, 10024, 2121, 6643, 8956, 9884, 8159, 6117 có độ chính tra cứu trung bình là 7%, các truy vấn có tỉ lệ chênh lệch giữa hai lớp thấp như 7548, 10356, 3775, 5103, 5925 có độ chính tra cứu trung bình là 73.2%. Điều này cũng có cho

¹ <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval> (Download lúc 6:32 AM ngày 25/12/2016)

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

khẳng định của chúng tôi rằng “hiệu năng của hệ thống tra cứu ảnh sử dụng SVM tăng cao khi số lượng mẫu của hai lớp trong tập ví dụ huấn luyện cân bằng nhau”.

Nhận định về sự cân bằng của hai lớp trong tập ví dụ huấn luyện làm tăng độ chính xác của tra cứu ảnh sử dụng SVM ở trên cũng là động lực chính của phương pháp của chúng tôi. Chúng tôi đề xuất một phương pháp tra cứu ảnh phản hồi liên quan hiệu quả RFBC (Image retrieval using relevance feedback with balanced classes), cho phép nâng cao độ chính xác hệ thống tra cứu ảnh thông qua việc cân bằng số mẫu trong mỗi lớp của tập ví dụ huấn luyện.

Phần còn lại bài báo này được tổ chức như sau: Trong phần 2, chúng tôi trình bày phương pháp tra cứu ảnh đề xuất. Phần 3 mô tả các thực nghiệm độ chính xác của chúng tôi và thảo luận các kết quả. Cuối cùng, chúng tôi đưa ra kết luận trong phần 4.

II. PHƯƠNG PHÁP TRA CỨU ẢNH ĐỀ XUẤT

Trong phần này, đầu tiên chúng tôi trình bày cơ sở lý thuyết được sử dụng trong phương pháp đề xuất. Sau đó chúng tôi trình bày chi tiết phương pháp đề xuất.

A. Cơ sở lý thuyết

1. Máy vectơ hỗ trợ

Đối với tập dữ liệu huấn luyện $D = \{x_1, x_2, \dots, x_n\}$, x_i là vectơ đặc trưng trong không gian R^m , m là số chiều của không gian, cùng các nhãn tương ứng của chúng $D_{\text{label}} = \{y_1, y_2, \dots, y_n\}$, $y_i \in \{-1, 1\}$. Mục đích là đi tìm một siêu phẳng phân tách có thể phân chia tập D sao cho tất cả các điểm có cùng nhãn nằm ở cùng một phía của siêu phẳng phân tách.

$$f(x) = w \cdot x + b = 0, w \in R^m, b \in R \quad (1)$$

ở đây x là một vectơ đầu vào, w là trọng số vectơ và b là độ lệch. Các siêu phẳng phân tách này là đường thẳng trong không gian hai chiều, là mặt phẳng trong không gian ba chiều, và trong trường hợp tổng quát hơn thì là không gian con R^{m-1} chiều trong không gian R^m chiều.

SVM cố gắng đi tìm hai tham số vectơ trọng số w và độ lệch b cho siêu phẳng phân tách tối ưu bằng cách cực đại lề hình học sao cho tất cả các điểm dữ liệu (x_i, y_i) đều thỏa mãn:

$$y_i(w \cdot x_i + b) \geq 1 \quad (2)$$

Các điểm vectơ hỗ trợ là những điểm có khoảng cách gần nhất tới siêu phẳng, bằng $1/\|w\|$ tức là $y_i f(x_i) = 1$, đại lượng $2/\|w\|$ được gọi là lề, và siêu phẳng phân tách tối ưu là siêu phẳng có giá trị lề lớn nhất hay còn gọi là siêu phẳng với lề cực đại. Như vậy việc tìm siêu phẳng tối ưu tức là đi tìm siêu phẳng có giá trị $\|w\|^2$ nhỏ nhất thỏa mãn (2)

$$\min h(w) = \frac{1}{2} \|w\|^2 \quad (3) \text{ thỏa mãn } y_i(w \cdot x_i + b) \geq 1, i = 1, \dots, n$$

Với $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ là các nhân tử Lagrange khác không, hàm Lagrangian tổng quát của bài toán tối là hàm

$$L(w, b, \alpha) = h(w) - \sum_{i=1}^n \alpha_i (y_i(w \cdot x_i + b) - 1) \quad (4)$$

$$\text{hay } L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i(w \cdot x_i + b) - 1) \quad (5)$$

Đặt đạo hàm của hàm số Lagrangian bằng không đối với các biến số w , và b , chúng ta có các mối quan hệ sau:

$$\frac{\partial}{\partial w} L(w, b, \alpha) = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \Leftrightarrow w = \sum_{i=1}^n \alpha_i y_i x_i \quad (6)$$

$$\frac{\partial}{\partial b} L(w, b, \alpha) = \sum_{i=1}^n \alpha_i y_i = 0 \quad (7)$$

Thay thế các quan hệ này vào $L(w, b, \alpha)$ chúng ta có hàm mục tiêu

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (8) \text{ thỏa mãn } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, n$$

Hàm quyết định cho phép phân lớp một mẫu z được cho bởi công thức:

$$\text{class}(z) = \text{sign}(f(z)) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i x_i \cdot z + b\right) \quad (9)$$

Khi dữ liệu là phân tách phi tuyến tính, SVM dùng hàm kernel

$$G(x, y) = \langle \varphi(x), \varphi(y) \rangle \quad (10)$$

Dữ liệu sẽ được ánh xạ sang một không gian mới bằng hàm φ với số chiều lớn hơn mà có thể phân tách tuyến tính. Hàm quyết định trở thành

$$\text{class}(z) = \text{sign}(f(z)) = \text{sign}\left(\sum_{i=1}^s \alpha_i y_i G(x_i, z) + b\right) \quad (11)$$

Trong đó $f(z)$ là đầu ra của hàm siêu phẳng quyết định của SVM và s là số lượng các điểm vectơ hỗ trợ. Hiệu quả của bộ phân lớp SVM phụ thuộc vào số lượng điểm vectơ hỗ trợ

2. Nhân RBF (Radial Basis Function Kernel)

Nhân RBF hoặc nhân Gaussian sẽ ánh xạ các vectơ sang một không gian mới nhiều chiều hơn thông qua hàm φ , tại đó cho ta một siêu phẳng tách tuyến tính với khoảng cách tối đa. RBF kernel được định nghĩa như sau:

$$G(x, y) = \langle \varphi(x), \varphi(y) \rangle = \exp(-\|x - y\|^2) \quad (12)$$

Hàm nhân sẽ tính toán tích vô hướng giữa hai vectơ. Nếu 2 vectơ mà tương tự nhau thì $\|x - y\|$ sẽ nhỏ, và làm cho $-\|x - y\|^2$ sẽ lớn, vì thế các vectơ gần nhau sẽ có giá trị hàm nhân RBF sẽ lớn hơn các vectơ xa nhau.

3. Khoảng cách từ một điểm (ảnh) tới siêu phẳng tách tối ưu

Để tính khoảng cách từ một điểm z tới siêu phẳng tách tối ưu (từ nay trở đi gọi là siêu phẳng tách), được dựa vào các vectơ hỗ trợ $S = \{x_1, x_2, \dots, x_k\}$.

$$\psi(z) = d(z, H) = \sum_{j=1}^k \alpha_j y_j x_j * x + b \quad (13)$$

y_j là nhân của các vectơ hỗ trợ tương ứng x_j ; α_j và b là các tham số được xác định tự động thông qua bộ phân lớp

4. Khoảng cách từ một ảnh tới một tập ảnh

Chúng tôi tính một cách đơn giản bằng cách tính khoảng cách trung bình của một thực thể z trong tập ảnh tới tất cả các thực thể z' còn lại nằm trong tập ảnh đó:

$$\xi(z) = \frac{1}{N} \sum_{i=1}^l \text{dist}(z, z'_i) \quad (14)$$

N là số lượng ảnh trong tập ảnh.

Hàm **dist** là khoảng cách giữa 2 ảnh thông qua hàm nhân theo công thức (12) giữa chúng:

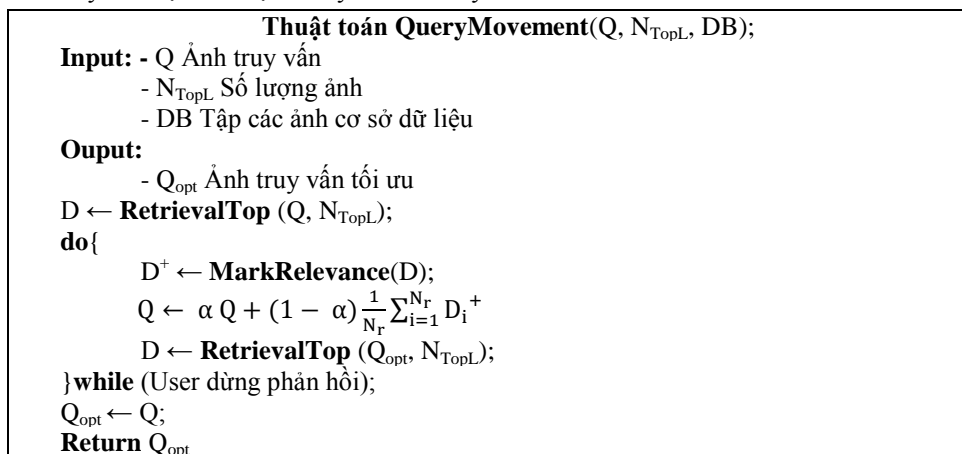
$$\text{dist}(z, z') = G(z, z') \quad (15)$$

B. Phương pháp đề xuất

1. Thuật toán xác định vị trí xây dựng tập ví dụ huấn luyện

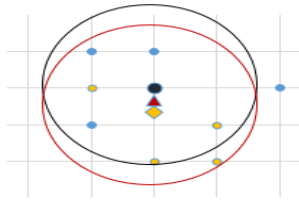
Các kỹ thuật tra cứu ảnh sử dụng SVM đã có xây dựng các lớp của tập ví dụ huấn luyện từ các điểm lân cận của điểm truy vấn trong không gian đặc trưng. Do đó, vị trí của điểm truy vấn sẽ ảnh hưởng rất lớn đến việc thu tập ví dụ huấn luyện. Xuất phát từ quan điểm này, chúng tôi sẽ thực hiện việc dịch chuyển điểm truy vấn. Ý tưởng của thuật toán dịch chuyển điểm truy vấn như sau: Khi người dùng gửi vào một truy vấn, thuật toán thực hiện tra cứu để được tập kết quả khởi tạo. Tiếp theo, dựa trên tập phản hồi liên quan mà người dùng phản hồi trên tập kết quả khởi tạo, thuật toán sẽ dịch chuyển điểm truy vấn đến vị trí gần các điểm liên quan. Quá trình trên được thực hiện cho đến khi thỏa mãn điều kiện đặt ra hoặc người dùng ngừng phản hồi.

Hình 1 dưới đây là thuật toán dịch chuyển điểm truy vấn



Hình 1. Thuật toán dịch chuyển điểm truy vấn **QueryMovement**

Thuật toán **QueryMovement** trên Hình 1 được thực hiện như sau: Ảnh Q_{opt} sẽ được dịch chuyển gần tới trọng tâm của các ảnh liên quan và được sử dụng như một truy vấn mới ở lần lặp tra cứu tiếp theo. Trọng số α nằm trong khoảng (0,1) thể hiện độ quan trọng của chúng. Các ảnh liên quan tại mỗi lần lặp được lựa chọn bởi người dùng giúp phản ánh chính xác hơn nhu cầu thông tin mà bản thân người dùng cần. Trong các lần lặp sau đó chỉ có truy vấn tối ưu mới được xem như là ảnh truy vấn mới đưa vào tra cứu, còn với các ảnh truy vấn tối ưu lần trước đó bị bỏ qua. Điều này có nghĩa là tầm quan trọng của các ảnh tích cực trước đó dần dần được tích hợp trong ảnh truy vấn tối ưu.



Hình 2. Ví dụ dịch chuyển điểm truy vấn **QueryMovement**

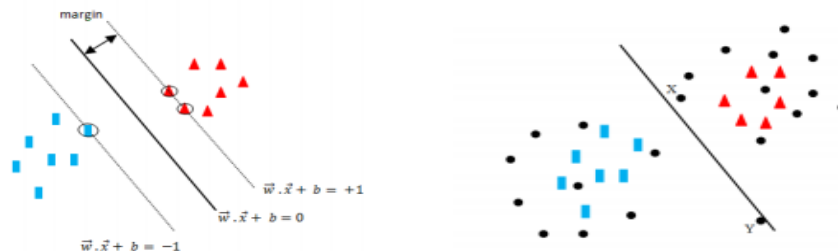
Ví dụ trên Hình 2, các điểm tròn là các ảnh trong không gian đặc trưng, điểm tròn đen là ảnh truy vấn. Với tra cứu truyền thống thu được các điểm nằm trong vòng tròn to màu đen, lựa chọn các ảnh liên quan với truy vấn Q là các điểm tròn màu vàng trong vòng tròn đen ta có được trọng tâm điểm hình thời màu vàng. Áp dụng thuật toán **QueryMovement** ta thu được truy vấn tối ưu là điểm tam giác màu đỏ, tra cứu với truy vấn tối ưu thu được các kết quả nằm trong vòng tròn màu đỏ.

2. Thuật toán bổ sung mẫu vào tập ví dụ huấn luyện

Tại pha tiếp theo chúng tôi sử dụng SVM như một bộ phân lớp những ảnh thu được sau khi thu hồi được từ ảnh truy vấn tối ưu cuối cùng. Thông qua SVM với những ảnh nào liên quan và không liên quan do người dùng xác định giúp ta có thể thu được tập ảnh có liên quan hơn cho nhu cầu thông tin của người dùng. Trong lần lặp đầu tiên, sử dụng bộ phân lớp SVM, sẽ cho ta một siêu phẳng tách tối ưu giữa ảnh liên quan và không liên quan trong toàn bộ ảnh csdl.

SVM hầu như được xem là một bộ phân lớp nhị phân tuyến tính, tuy nhiên nó cũng có thể thực hiện hiệu quả phân loại phi tuyến tính bằng cách ánh xạ các đối tượng từ không gian ban đầu vào một không gian đặc trưng có số chiều cao hơn mà tại đó có thể phân tách tuyến tính.

Ý tưởng chính việc áp dụng SVM là để tìm được một siêu phẳng tách tối ưu w , sẽ phân tách dữ liệu huấn luyện theo cách tối ưu hóa cực đại lề giữa hai lớp. Những điểm vectơ hỗ trợ là những điểm có khoảng cách gần biên quyết định nhất như dưới Hình 3 minh họa là các điểm được khoanh tròn.



Hình 3. a. Bộ phân lớp tuyến tính nhị phân SVM **b.** Bộ phân lớp tuyến tính nhị phân SVM trong toàn bộ dữ liệu

Như trong Hình 3a, hình tam giác, hình chữ nhật tương ứng với các dữ liệu huấn luyện ban đầu thuộc hai lớp là tích cực (+1) và tiêu cực (-1). Dữ liệu được khoanh tròn nằm trên đường biên của lề là các vectơ hỗ trợ và còn đường màu đen đậm được gọi là biên quyết định.

Các điểm nằm trong lề của cả 2 lớp (hay gọi là trong ống) là những mẫu sẽ cung cấp cho ta những cái nhìn sâu sắc nhất trong phân phối dữ liệu. Như trên Hình 3b trong số các điểm chưa được gán nhãn khác nhau thì có 3 ứng cử viên nằm trong ống. Nhìn một cách trực quan chúng ta có thể thấy được nếu yêu cầu gợi ý nhãn của mẫu gần nhất với ranh giới biên quyết định thì sẽ cho ta lựa chọn tốt. Nhãn của các ứng cử viên khác không nằm trong ống dường như rõ ràng phù hợp với các lớp của các mẫu bên tương ứng, hoặc có thể làm sai lệch đi kết quả phân lớp đối với các mẫu đã được gán nhãn. Mô hình học sẽ được cải thiện khi việc gợi ý nhãn này là do người dùng xác định cho phù hợp với nội dung ngữ nghĩa mà đang muốn tra cứu. Những ảnh không chắc chắn nhất là những ảnh gần với biên quyết định nhất. Chúng tôi tính khoảng cách ψ của từng ảnh nằm trong hai lề tới siêu phẳng theo công thức (13).

Nhưng đôi khi chúng ta chú ý quá nhiều đến một số khu vực nhất định của không gian đặc trưng và loại bỏ các khu vực khác đại diện cho phân phối cơ bản. Điều này dẫn đến khi chọn những điểm gần ranh giới biên quyết định cho ta kết quả không tốt với những điểm xa hơn. Như trong Hình 3b chúng ta thấy một mẫu như thế có thể gây hiểu nhầm cho phân loại. Mặc dù Y gần với siêu phẳng tách nhất, nhưng nó cách xa các mẫu khác trong không gian đặc trưng và nó có thể không đại diện cho phân phối cơ bản. Trong trường hợp này chọn mẫu X là khả thi hơn bởi vì nó gần với các mẫu khác, nơi có sự phân loại đúng để đạt được một mô hình chính xác.

Mẫu như X chúng tôi gọi là mẫu có tính đại diện hơn. Khoảng cách ξ từ một mẫu tới các mẫu còn lại nằm trong hai lề của siêu phẳng phân tách tính theo công thức (14). Quá trình chọn ảnh nào để đưa tiếp vào cho tập huấn luyện, chúng tôi dựa vào sự kết hợp cả tính không chắc chắn và tính đại diện của ảnh được lựa chọn:

$$\hat{z} = argmin(\gamma * \psi(z) + (1 - \gamma) * \xi(z))$$

Hình 4 dưới đây là thuật toán bổ sung mẫu vào tập ví dụ huấn luyện.

Thuật toán InsertTraningSet(Q, N_{TopL}, DB);

Input:

- D Tập các ảnh được gán nhãn
- D_{label} Nhãn tương ứng của tập X
- DB Tập các ảnh toàn bộ cơ sở dữ liệu

Ouput:

- \hat{z} ảnh bổ sung tập huấn luyện

```

< $\alpha, b, SV, SV^{label}$ >  $\leftarrow$  Classifier(D, Dlabel);
for i  $\leftarrow$  1 to DB.Count do
   $\psi(DB_i) \leftarrow \sum_{j=1}^k \alpha_j SV_j^{label} SV_j * DB_i + b$ 
  if ( $|\psi(DB_i)| \leq margin$ ) {
    X  $\leftarrow$  X  $\cup$  DBi;  $\psi(X) \leftarrow \psi(X) \cup \psi(DB_i)$ ;
  }
for i  $\leftarrow$  1 to X.Count do
  for j  $\leftarrow$  1 to X.Count do
     $\xi' \leftarrow \xi' + dist(X_i, X_j)$ 
   $\xi(X_i) \leftarrow \frac{1}{X.Count} \xi'$ 
for i = 1 to X.Count do
   $\mu(X_i) = \gamma * \psi(X_i) + (1 - \gamma) * \xi(X_i)$ 
 $\hat{z} = argmin(\mu(X))$ ;
Return  $\hat{z}$ ;

```

Hình 4. Thuật toán bổ sung mẫu vào tập huấn luyện

Thuật toán **InsertTraningSet** trên Hình 4 thực hiện như sau:

Với đầu vào là một danh sách các ảnh D_1, D_2, \dots, D_n cùng nhãn D_{label} tương ứng của tập D thuật toán sẽ lấy được các tham số $\langle \alpha, b \rangle$ và các vectơ hỗ trợ SV cũng như nhãn tương ứng SV^{label} . Thuật toán duyệt từng ảnh trong toàn bộ cơ sở dữ liệu DB để tính khoảng cách từng ảnh tới siêu phẳng phân tách lấy ra tập X các ảnh nằm trong lề và khoảng cách $\psi(X)$ của các ảnh đó. Tiếp chúng tôi tính khoảng cách của từng điểm trong tập X tới các điểm còn lại thu được khoảng cách $\xi(X)$ sau đó kết hợp hai khoảng cách này với trọng số γ nằm trong khoảng (0,1) thể hiện độ quan trọng của chúng thu được $\mu(X)$. Dựa vào đó thuật toán chọn được một ảnh \hat{z} vừa gần siêu phẳng phân tách vừa gần các điểm còn lại nằm trong hai lề của siêu phẳng.

3. Thuật toán tra cứu ảnh phản hồi liên quan

Như vậy, đến đây, chúng ta đã tìm được vị trí của điểm truy vấn để xây dựng tập ví dụ huấn luyện cân bằng với thuật toán **QueryMovement** và bổ sung mẫu vào tập ví dụ huấn luyện qua thuật toán **InsertTraningSet**. Do đó, chúng ta có thể xây dựng được thuật toán tra cứu ảnh để thực hiện tra cứu ảnh.

Hình 5 dưới đây là thuật toán tra cứu ảnh phản hồi liên quan RFBC:

Thuật toán RFBC

Input: -Q Ảnh truy vấn
-N_{TopL} Số lượng ảnh
-DB Tập các ảnh cơ sở dữ liệu

Ouput: -D Tập ảnh kết quả

```

Qopt  $\leftarrow$  QueryMovement (Q, NTopL, DB);
D  $\leftarrow$  RetrievalTop (Qopt, NTopL, DB);
D+  $\leftarrow$  MarkRelevance(D);  $D_{label}^+ \leftarrow$  SetLabel(D+, 1);
D-  $\leftarrow$  MarkIrrelevance(D);  $D_{label}^- \leftarrow$  SetLabel(D-, -1);
X  $\leftarrow$  D+  $\cup$  D-;  $X_{label} \leftarrow$  D+label  $\cup$  D-label;
< $\alpha, b, SV, SV^{label}$ >  $\leftarrow$  Classifier(X, Xlabel);
<X, Xlabel>  $\leftarrow$  PointOptimal(DB, SV);
X  $\leftarrow$  Union(X, x);  $X_{label} \leftarrow$  Union(Xlabel, xlabel);
< $\alpha, b, SV, SV^{label}$ >  $\leftarrow$  Classifier(X, Y);
for i  $\leftarrow$  1 to DB.Count do
   $\psi(DB_i) \leftarrow \sum_{j=1}^k \alpha_j SV_j^{label} SV_j * DB_i + b$ 
D  $\leftarrow$  Ranking(DB,  $\psi$ , NTopL);

```

Hình 5. Thuật toán tra cứu ảnh RFBC

Thuật toán RFBC trên Hình 5 được thực hiện như sau:

Khi người dùng gửi vào một truy vấn trên giao diện truy vấn bởi mẫu, thuật toán sẽ thu được ảnh truy vấn tối ưu Q_{opt} thông qua thuật toán **QueryMovement** (xem thuật toán **QueryMovement** trên hình 1). Với ảnh truy vấn mới Q_{opt} tối ưu này thuật toán sẽ lấy được tập ảnh D gồm N_{TopL} ảnh đầu tiên được phân hạng đầu tiên trong toàn bộ cơ sở dữ liệu DB thông qua hàm **RetrievalTop()**. Tiếp theo các ảnh trong tập D sẽ được người dùng gắn nhãn lựa chọn ảnh nào liên quan mang nhãn +1, không liên quan mang nhãn -1 thông qua các hàm **MarkRelevance**, **MarkIrrelevance** và **SetLabel**. Gộp cả hai tập liên quan và không liên quan này thu được tập huấn luyện X cùng nhãn tương ứng X_{label} của chúng đưa vào tiến hành huấn luyện phân lớp sử dụng hàm **Classifier()** dùng SVM được mô tả ở phần đầu nội dung 2. Sau khi phân lớp chúng tôi lựa chọn được một ảnh z dựa vào thuật toán **InsertTrainingSet** (xem thuật toán **InsertTrainingSet** trên hình 4). Tập huấn luyện mới lúc này bao gồm cả ảnh z vừa tìm được sẽ đưa vào huấn luyện phân lớp lại thu được siêu phẳng phân tách tối ưu mới. Thuật toán phân hạng lại các ảnh trong toàn bộ cơ sở dữ liệu dựa vào giảm dần khoảng cách ψ từng ảnh trong cơ sở dữ liệu DB tới siêu phẳng phân tách mới này thông qua hàm **Ranking()** và thu được tập ảnh trả về cuối cùng D gồm N_{TopL} ảnh phân hạng đầu tiên.

III. ĐÁNH GIÁ THỰC NGHIỆM

A. Môi trường thực nghiệm

1. Cơ sở dữ liệu ảnh

Trong thực nghiệm, chúng tôi sử dụng một tập con của Corel Photo Gallery làm cơ sở dữ liệu thử nghiệm để đánh giá hiệu quả của phương pháp. Tập này gồm 80 loại, ví dụ như là: mùa thu, cây cảnh, lâu đài, voi, linh trưởng, hỏa tiễn, hồ, tàu hỏa, thác nước. Tất cả các ảnh trong tập ảnh này có tính chất là đều chứa đối tượng tiền cảnh nổi bật. Đa số nhóm đều gồm 100 ảnh, có một vài nhóm có hơn 100 hình ảnh. Cỡ của các ảnh có $\max(\text{chiều rộng, chiều cao}) = 120$ và $\min(\text{chiều rộng, chiều cao}) = 80$.

2. Vectơ đặc trưng

Đặc trưng chúng tôi sử dụng gồm hai loại đặc trưng: màu và kết cấu (xem Bảng 3 ở dưới).

Bảng 3. Các loại đặc trưng

Các loại đặc trưng		Tên đặc trưng	Độ dài
Loại đặc trưng màu	Lược đồ màu	hsvHistogram	32
	Tương quan màu	color auto correlogram	64
	Gắn kết màu	colorMoments	6
Loại đặc trưng kết cấu	Biến đổi wavelet	waveletTransform	40
	gabor Wavelet	gaborWavelet	48

Đối với đặc trưng màu chúng tôi trích rút 2 đặc trưng gắn kết màu: trung bình, độ lệch chuẩn trong mỗi kênh màu RGB có độ dài 6-D. Sau đó 32-D (8x2x2) lược đồ màu HSV được tính, kênh hue được lượng hóa thành 8 bin, cả hai kênh S và V đều được lượng hóa thành 2 bin. Đặc trưng màu cuối cùng là tương quan màu độ dài 64-D (4x4x4) trong không gian RGB.

Chúng tôi trích rút đặc trưng kết cấu: đặc trưng gabor gồm Mean-squared energy và meanAmplitude cho 4 scale, 6 hướng đối với ảnh đa cấp xám; đặc trưng gắn kết wavelet 48-D gồm hai gắn kết trung bình, độ lệch chuẩn.

3. Biểu diễn ảnh

Mỗi ảnh được sử dụng biểu diễn bởi năm đặc trưng trực quan gồm ba đặc trưng màu và hai đặc trưng kết cấu. Các vectơ đặc trưng tương ứng với mỗi kênh là một bảng hai chiều gồm 10800 dòng (mỗi dòng chứa một vectơ đặc trưng của ảnh) và 190 cột (độ dài tổng của một vectơ đặc trưng).

4. 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 1981320 dòng.

B. 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, một truy vấn khởi tạo sẽ được thực hiện để tạo ra kết quả truy vấn khởi tạo. Tiếp theo chúng tôi mô phỏng tương

tác người dùng bằng việc chọn 100 ảnh liên quan từ kết quả tra cứu khởi tạo dựa vào tập tin cây nền. Những ảnh liên quan từ việc phản hồi sẽ được dùng để xác định được vị trí tối ưu của truy vấn dựa vào thuật toán **QueryMovement** để hình thành truy vấn tiếp theo. Quá trình này có thể sẽ dừng lại khi người dùng không muốn tiếp tục nhưng trong thực nghiệm chúng tôi thực hiện theo hai kịch bản lặp lại một lần và lặp lại năm lần.

Sau khi có kết quả với pha xác định truy vấn tối ưu chúng ta thu được 100 ảnh đầu tiên gần nhất so với ảnh truy vấn Q. Lúc này chúng tôi chọn các ảnh liên quan trong 100 ảnh đó cũng dựa vào tập tin cây nền nghĩa là chúng có cùng khái niệm ngữ nghĩa với ảnh truy vấn hay không? Sau đó hai tập ảnh liên quan và không liên quan được tạo ra: các ảnh được chọn mang nhãn +1 còn những ảnh còn lại sẽ được gán nhãn -1. Hai tập ảnh này được chúng tôi sử dụng làm tập huấn luyện cho bộ phân lớp được nêu tại mục “máy vectơ hỗ trợ” tại phần đầu mục II, kết quả trả về cho ta được một siêu phẳng phân tách các ảnh trong tập huấn luyện. Trong thực nghiệm chúng tôi cũng xây dựng hệ thống thu lấy kết quả độ chính xác tra cứu ảnh khi thực hiện phân lớp SVM truyền thống trên tập huấn luyện thu được từ tra cứu truyền thống gọi tên là **SVMRF**.

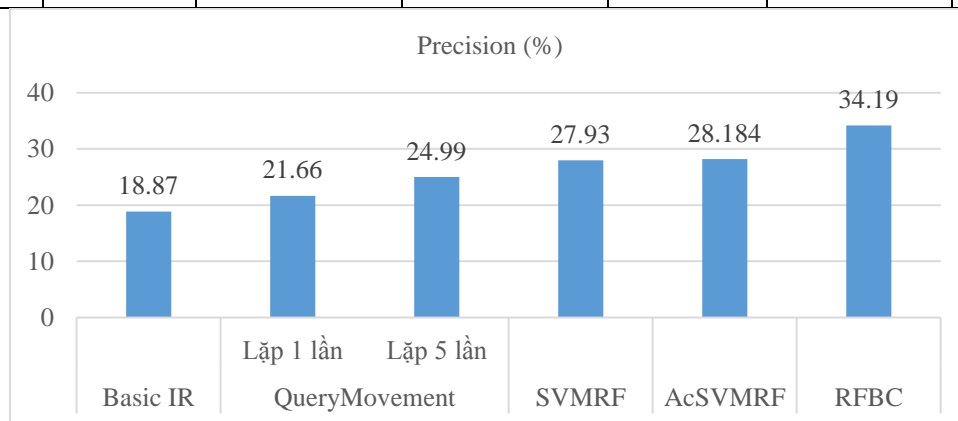
Dùng siêu phẳng phân tách này chúng tôi tìm được một ảnh z trong những ảnh nằm trong ống mà vừa gần với siêu phẳng phân tách lại vừa gần nhất với các ảnh còn lại nằm trong ống. Nhãn của ảnh này được gán dựa vào tập tin cây nền, thêm vào tập huấn luyện ban đầu, bây giờ tập huấn luyện ban đầu gồm 101 ảnh được đưa vào huấn luyện lại với bộ phân lớp cho ta một siêu phẳng phân tách tối ưu mới. Sắp xếp lại toàn bộ ảnh trong cơ sở dữ liệu theo khoảng cách của từng ảnh tới siêu phẳng ta thu được top 100 ảnh cuối cùng. Kết quả thu được cuối cùng chính là kết quả độ chính xác của phương pháp **RFBC** chúng tôi đề xuất với năm lần lặp để định vị truy vấn tối ưu. Bên cạnh đó kết quả thực nghiệm tra cứu ảnh khi huấn luyện bộ phân lớp SVM tích cực vừa nêu với tập huấn luyện thu được từ tra cứu truyền thống cũng được lưu lại để so sánh, đánh giá gọi là **AcSVMRF**.

Tất cả 10800 ảnh trong tập ảnh được dùng làm các truy vấn. Độ chính xác trung bình ở mức 100 ảnh trả về được sử dụng để đánh giá. Bốn phương pháp khác nhau được sử dụng để so sánh bao gồm **Basic IR** (hệ thống CBIR truyền thống), **QueryMovement** (1 lần lặp và 5 lần lặp), **SVMRF**, **AcSVMRF** với hệ thống **RFBC** mà chúng tôi đề xuất.

Có nhiều chỉ số đánh giá khác nhau được đề xuất để đánh giá hiệu quả của các hệ thống CBIR, chúng tôi sử dụng độ đo cơ bản là độ chính xác (thực nghiệm là 100 ảnh trả về). Các kết quả, độ chính xác trung bình của 10800 truy vấn, được thể hiện bằng số liệu trong Bảng 4 và bằng đồ thị trong Hình 6 ở dưới. Do giới hạn về không gian bài báo, chúng tôi chỉ trình bày trong bài báo này độ chính xác trung bình của từng phương pháp còn chi tiết về độ chính xác trung bình của từng 80 loại truy vấn xem tại địa chỉ <http://117.6.134.238:368/results/RFBC.html>.

Bảng 4. Bảng kết quả của 5 phương pháp

Phương pháp	Basic IR	QueryMovement		SVMRF	AcSVMRF	RFBC
		Lặp 1 lần	Lặp 5 lần			
Precision (%)	18.87	21.66	24.99	27.93	28.184	34.19



Hình 6. So sánh độ chính xác

Nhìn vào kết quả thực nghiệm trên Hình 6 chúng tôi có thể đưa ra các kết luận: Độ chính xác của phương pháp đề xuất tăng lên 81% so với phương pháp tra cứu truyền thống. Số lần lặp xác định truy vấn tối ưu tăng lên cũng làm cho hiệu quả tăng 3.33% từ 1 lần lặp là 21.66% lên 5 lần lặp là 24.99%. Độ chính xác phương pháp chúng tôi cũng tăng đáng kể xấp xỉ 6% so với độ chính xác khi chỉ phân lớp trên kết quả của tra cứu truyền thống.

IV. KẾT LUẬN

Chúng tôi đã đề xuất phương pháp tra cứu ảnh dựa vào nội dung cho cải tiến độ chính xác tra cứu của các hệ thống tra cứu phản hồi liên quan sử dụng SVM truyền thống. Phương pháp của chúng tôi quan tâm đến việc thu được hai lớp có số lượng mẫu cân bằng nhau. Để thu được sự cân bằng, chúng tôi thực hiện tận dụng thông tin phản hồi của

người dùng tại mỗi lần lặp để dịch chuyển điểm truy vấn đến nơi mà lân cận của nó có thể giúp tạo ra hai lớp có số lượng mẫu cân bằng nhau. Trên cơ sở hai lớp có số lượng mẫu cân bằng, phương pháp của chúng tôi sử dụng SVM tích cực để phân lớp các ảnh liên quan ngữ nghĩa và phân hạng các ảnh theo thứ tự giảm dần của khoảng cách từ một điểm đến siêu phẳng tách tối ưu.

Kết quả thực nghiệm của chúng tôi trên cơ sở dữ liệu đặc trưng gồm 10800 ảnh đã chỉ ra rằng phương pháp được đề xuất **RFBC** cung cấp một độ chính xác cao hơn hẳn so với các phương pháp **Basic IR** (hệ thống CBIR truyền thống), **QueryMovement**, **SVMRF**, **AcSVMRF**.

LỜI CẢM ƠN

Chúng tôi xin cảm ơn đề tài "Nghiên cứu phương pháp tra cứu ảnh dựa vào đa truy vấn", Mã số: PTNTĐ17.04, thuộc Phòng thí nghiệm trọng điểm Công nghệ mạng và Đa phương tiện, Viện CNTT, Viện HLKHCNVN đã hỗ trợ.

TÀI LIỆU THAM KHẢO

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349 -1380, Dec. 2000.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1-60, May 2008.
- [3] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1088 -1099, July 2006.
- [4] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 716 -727, 2006.
- [5] L. Liu, M. Yu, and L. Shao, "Multiview alignment hashing for efficient image search," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 956-966, 2015.
- [6] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 1019-1034, May 2015.
- [7] Y. Rui, T.S. Huang, Ortega M and Mehrotra S. "Relevance feedback: A power tool in interactive contentbased image retrieval". *IEEE Tran. Circuits and Systems for Video Tech.* 8(5): 644-655, Sep. 1998.
- [8] X. Zhou and T. Huang, "Relevance feedback for image retrieval: A comprehensive review," *Multimedia Syst.*, vol. 8, no. 6, pp. 536-544, Apr. 2003.
- [9] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Process.*, 1997, vol. 2, pp. 815-818.
- [10] T. S. Huang and X. S. Zhou, "Image retrieval by relevance feedback: From heuristic weight adjustment to optimal learning methods," in *Proc. IEEE ICIP, Thessaloniki, Greece*, Oct. 2001, pp. 2-5.
- [11] J. Laaksonen, M. Koskela, and E. Oja, "PicSOM: Self-organizing maps for content-based image retrieval," in *Proc. IJCNN*, Washington, DC, 1999, pp. 2470-2473.
- [12] Y. Chen, X.-S. Zhou, and T.-S. Huang, "One-class SVM for learning in image retrieval," in *Proc. IEEE ICIP*, 2001, pp. 815-818.
- [13] C. Hoi, C. Chan, K. Huang, M. R. Lyu, and I. King, "Biased support vector machine for relevance feedback in image retrieval," in *Proc. IJCNN*, Budapest, Hungary, Jul. 25-29, 2004, pp. 3189-3194.
- [14] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 189-201, Feb. 2008.
- [15] W. Bian and D. Tao, "Biased discriminant Euclidean embedding for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 545-554, Feb. 2010.
- [16] X. Zhou and T. Huang, "Small sample learning during multimedia retrieval using BiasMap," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. 11-17.
- [17] D. Tao, X. Tang, X. Li, and Y. Rui, "Kernel direct biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716-724, Aug.
- [18] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang, "Marginal Fisher Analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811-2821, Nov. 2007.
- [19] D. Cai, X. He, and J. Han, "Regularized regression on image manifold for retrieval," in *Proc. 9th ACM SIGMM Int. Workshop MIR*, Augsburg, Germany, Sep. 2007, pp. 11-20.

- [20] P. Hong, Q. Tian, and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," in Proc. IEEE ICIP, Vancouver, BC, Canada, 2000, pp. 750-753.
- [21] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 7, pp. 1088-1099, Jul. 2007.
- [22] G. Guo, A. K. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," IEEE Trans. Neural Netw., vol. 13, no. 4, pp. 811-820, Jul. 2002.
- [23] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
- [24] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in Proc. ACM Multimedia, 2001, pp. 107-118.
- [25] J. Li, N. Allinson, D. Tao, and X. Li, "Multitraining support vector machine for image retrieval," IEEE Trans. Image Process., vol. 15, no. 11, pp. 3597-3601, Nov. 2006.
- [26] L. Zhang, F. Lin, B. Zhang, "Support Vector Machine Learning For Image Retrieval," in Image Processing, 2001. Proceedings. 2001 International Conference, pp 721 - 724 vol.2, oct 2001.

AN IMAGE RETRIEVAL METHOD EFFICIENTLY EXPLOIT USER'S FEEDBACK

Cu Viet Dung, Nguyen Huu Quynh, Ngo Quoc Tao, An Hong Son, Le Manh Hung

ABSTRACT: *There have been many methods image retrieval with relevance feedback using Support Vector Machines SVM. However, these methods have not interested to the balance of the two classes in the user's feedback process so low accuracy. In this paper, we propose an effective image retrieval with relevance feedback method, called RFBC (Image Retrieval Relevance Feedback with Balanced Classes), which improves the accuracy of image retrieval systems through balancing the number of samples in each class of the training example set. Our method utilizes the evaluation of the user to determine the position where two layers of balanced samples can be obtained. . We also provided empirical results on a database of 10,800 images to show the accuracy of the method..*

Keywords: *Content based image retrieval, feature space, optimal query, relevance feedback with balanced classes.*