

PHÉP CHUẨN HÓA DỮ LIỆU ĐẶC TRƯNG MỨC THẤP CỦA ẢNH TRONG TRUY VẤN ẢNH DỰA TRÊN NỘI DUNG

Hoàng Xuân Trung¹, Đoàn Văn Hòa², Nguyễn Tân Ân³, Ngô Hoàng Huy⁴

¹ Đại học Kinh doanh và Công nghệ Hà Nội

² Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Quân sự

³ Khoa Công nghệ Thông tin, Học viện Quản lý Giáo dục

⁴ Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

trungvniit@gmail.com, doanvanhoa@gmail.com, nguyentanan@yahoo.com, huyngo3i@gmail.com

TÓM TẮT: Trong vấn đề tra cứu ảnh dựa trên nội dung (CBIR), ảnh được biểu diễn bằng nhiều đặc trưng mức thấp để mô tả màu sắc, kết cấu và hình dạng của ảnh. Các thuật toán truy vấn ảnh thường xây dựng các độ đo tương tự toàn cục giữa các vectơ đặc trưng biểu diễn đối tượng ảnh. Sự kết hợp nhiều đặc trưng ảnh khác nhau như vậy đòi hỏi phải chuẩn hóa dữ liệu các bộ đặc trưng. Bước chuẩn hóa dữ liệu đặc trưng ảnh có ảnh hưởng đến chất lượng CBIR và đã có một số phép chuẩn hóa được sử dụng như chuẩn hóa min-max, chuẩn hóa 3-sigma và chuẩn hóa mở rộng 3sigma-FCM, ... Bài báo này trình bày một phương pháp chuẩn hóa mới cải tiến từ chuẩn hóa 3sigma-FCM. Thử nghiệm đã chứng tỏ tính hiệu quả của thuật toán đề xuất cho vấn đề chuẩn hóa dữ liệu đặc trưng ảnh. Khoảng động của dữ liệu đặc trưng ảnh được chuẩn hóa trên đoạn $[-1,1]$ rộng hơn khoảng động tương ứng của phép chuẩn hóa 3sigma-FCM. Thử nghiệm cũng chứng tỏ, phép chuẩn hóa mới cũng tăng chất lượng CBIR khi kết hợp với các thuật toán xây dựng độ đo tương tự ảnh như EMR.

Từ khóa: Tra cứu ảnh dựa trên nội dung, Đặc trưng mức thấp, FCM, Chuẩn hóa 3sigma-FCM, Chuẩn hóa tối ưu, EMR.

I. GIỚI THIỆU

Truy vấn ảnh dựa vào nội dung (CBIR) trở thành lĩnh vực nghiên cứu quan trọng trong những năm qua. Các hệ thống này thường trích rút các biểu diễn trực quan của ảnh và định nghĩa các hàm đo độ tương tự, tìm kiếm, so khớp mối liên quan để tra cứu theo yêu cầu.

Tuy nhiên việc truy vấn ảnh dựa trên nội dung đòi hỏi phải kết hợp các thông tin mô tả về màu sắc, kết cấu và hình dạng đồng thời. Mỗi vectơ đặc trưng trực quan trích rút được từ một ảnh có các thành phần giá trị số có thể thuộc các khoảng số rất khác nhau, do đó khi kết hợp nhiều đặc trưng biểu diễn cho các ảnh để xây dựng độ đo tương tự của một cặp ảnh chúng ta cần phải chuẩn hóa dữ liệu đặc trưng về một miền số chung, cụ thể là đoạn $[0, 1]$ (xem [1, 2]).

Mới gần đây, trong [3] các tác giả đã đề xuất một phương pháp chuẩn hóa gọi là 3sigma-FCM. Phương pháp chuẩn hóa này là một mở rộng của phương pháp chuẩn hóa 3sigma thông dụng, được sử dụng nhiều trong lĩnh vực CBIR [2, 4, 5, 6]. Phương pháp 3sigma-FCM áp dụng cho hệ thống tra cứu ảnh theo nội dung sử dụng phản hồi liên quan và kết hợp nhiều đặc trưng trực quan mức thấp khá hiệu quả. Tuy nhiên do khi kết nhập các giá trị chuẩn hóa 3sigma tính được dựa trên thuật toán FCM các tác giả chỉ dùng hệ số chia tỷ lệ theo tham số là số lớp nên khoảng động dữ liệu đặc trưng sau khi phân cụm vẫn thu hẹp hơn so với đoạn $[-1,1]$. Điều này làm giảm tính khái quát của phương pháp chuẩn hóa 3sigma-FCM để áp dụng phép chuẩn hóa này sang miền ứng dụng khác vì thường các bước chuẩn hóa dữ liệu đều yêu cầu trải rộng trên toàn bộ đoạn được quy chuẩn, chẳng hạn trong các phép xây dựng độ tương tự sử dụng nhiều bộ đặc trưng dựa trên tích phân Choquet [7].

Phương pháp của chúng tôi là khi kết nhập các giá trị được chuẩn hóa riêng rẽ cho từng cụm sử dụng một biến đổi tuyến tính với các hệ số tuyến tính được ước lượng tự động và tối ưu từ cơ sở dữ liệu đặc trưng ảnh. Những khác biệt này được trình bày trong phần ba và phần bốn.

Phần còn lại của bài báo được tổ chức như sau. Phần hai, một số nghiên cứu liên quan sử dụng kết hợp đặc trưng, chuẩn hoá đặc trưng, chuẩn hoá khoảng cách. Phần ba là đề xuất chuẩn hoá đặc trưng cải tiến của chúng tôi. Các kết quả thực nghiệm đưa ra trong phần bốn. Kết luận và hướng nghiên cứu tương lai trong phần năm.

II. NGHIÊN CỨU LIÊN QUAN

Bảng 1 sau đây liệt kê một số kí hiệu được sử dụng trong bài báo này.

Bảng 1. Các ký hiệu và các định nghĩa của nó

Kí hiệu	Định nghĩa
I	Ảnh nói chung
Q	Ảnh truy vấn
n	Kích thước của cơ sở dữ liệu ảnh
F	Không gian đặc trưng
m	Độ dài của một véc tơ đặc trưng
$T=\{t_1, \dots, t_k\}$	Tập kết hợp nhiều đặc trưng

Kí hiệu	Định nghĩa
E	Một điểm dữ liệu (đặc trưng)
E_i	Đặc trưng thô của ảnh thứ i
$E_{t,i}$	Đặc trưng thô bộ t của ảnh thứ i
E_i^{Norm}	Đặc trưng chuẩn hoá ảnh thứ i
$E_{t,i}^{Norm}$	Đặc trưng bộ t ảnh thứ i đã chuẩn hoá
Q_t	Dữ liệu bộ t ảnh truy vấn
Q_t^{Norm}	Dữ liệu chuẩn hoá bộ t của ảnh truy vấn
$V_{t,c,j}$	Tâm cụm c của thành phần đặc trưng j ở bộ đặc trưng t (theo FCM)
$\eta_{t,c,i}$	Giá trị độ thuộc của phần tử đặc trưng i ở bộ t so với tâm c, p là hệ số FCM
$\sigma_{t,c,j}$	Độ lệch chuẩn thành phần j của đặc trưng t theo cụm c

A. Các đặc trưng mức thấp trong CBIR và biểu diễn đối tượng ảnh

Trong các ứng dụng CBIR khác nhau, tập dữ liệu khác nhau yêu cầu phải có cách kết hợp các bộ đặc trưng phù hợp. Các đặc trưng trực quan được sử dụng là mô tả màu, kết cấu và hình dạng, xem chẳng hạn [8-13, 2, 14-17]. Dưới đây chúng ta liệt kê một số đặc trưng mức thấp cụ thể.

Đặc trưng màu (color features)

Color Moments: Mỗi ảnh được chia thành các vùng 3×3 và ba moment màu color mean, color variance and color skewness được tính toán cho mỗi vùng trong mỗi kênh màu HSV tương ứng [11].

Đặc trưng kết cấu (texture features)

Gabor Wavelets Texture: Mỗi ảnh được biểu diễn bởi 40 ảnh con bởi áp dụng biến đổi Gabor với 5 mức tỉ lệ và 8 hướng. Sau đó ba moment sẽ được tính toán cho mỗi ảnh con [17].

Local Binary Pattern (LBP): Mỗi điểm ảnh được gán nhãn dựa vào các láng giềng của nó trong cửa sổ 3×3 và được biểu diễn bởi một số nhị phân. Histogram của 256 (28) nhãn này được định nghĩa như là độ đo bất biến kết cấu [18].

Đặc trưng hình dạng (shape features)

Edge: Với đặc trưng này, biểu đồ hướng biên được sử dụng. Thông tin biên chứa trong ảnh được tạo ra và xử lý sử dụng thuật toán phát hiện biên Canny trên ảnh đã được chuyển đổi thành ảnh xám. Biểu đồ hướng biên sau đó được lượng tử hóa thành 36 bin với 10 độ cho mỗi bin. Một bin được sử dụng bổ sung để đếm số lượng các điểm ảnh không có thông tin cạnh [18].

GIST (Global Scene): là mô một mô tả tổng hợp thông tin gradient cho các phần khác nhau của ảnh. Với việc nhân chập ảnh với 32 bộ lọc Gabor tại 4 tỉ lệ và 8 hướng, 32 bản đồ đặc trưng cùng cỡ với ảnh gốc được tạo ra. Mỗi bản đồ đặc trưng này sau đó được chia thành 16 vùng bởi lưới 4×4 và giá trị đặc trưng trung bình của mỗi vùng được tính toán [19, 18].

Khái quát lại, trong [2] các tác giả đã trình bày một biểu diễn của ảnh sử dụng nhiều bộ đặc trưng, trong mỗi bộ đặc trưng cũng có nhiều thành phần như sau:

Định nghĩa 1 (Biểu diễn ảnh): $O \equiv O E, F, T$, trong đó:

(1) E: dữ liệu thô của ảnh

(2) $F = f_i$: tập đặc trưng mức thấp (mô tả về màu, kết cấu, hình dạng).

(3) $T = t_{ij}$: tập biểu diễn cho đặc trưng f_i , $t_{ij} \equiv [t_{ij1}, \dots, t_{ijk}, \dots, t_{ijk}]$ véc tơ gồm nhiều thành phần.

Sau khi đã biểu diễn được một ảnh, chúng ta sẽ cần xây dựng các độ đo tương tự của một cặp ảnh, có nhiều cách xây dựng độ đo tương tự, xem chẳng hạn [7, 20].

Đơn giản nhất để xây dựng một độ đo tương tự là xây dựng một độ đo khoảng cách với giá trị số trong đoạn $[0,1]$. Chẳng hạn, độ đo khoảng cách có trọng số trong [3] sau:

Độ đo khoảng cách D giữa hai đối tượng $O^1(E^1, F^1, T^1)$ và $O^2(E^2, F^2, T^2)$ dạng đầy đủ xác định:

(i) Độ đo khoảng cách một bộ đặc trưng:

$$D_{ij} \quad t_{ij}^1, t_{ij}^2 \stackrel{\text{def}}{=} D_{ij} \quad t_{ij}^1, t_{ij}^2, w_{ijk}$$

(ii) Độ đo khoảng cách toàn bộ:

$$D \quad O^1, E^1, F^1, T^1, O^2, E^2, F^2, T^2 \stackrel{\text{def}}{=} \sum_{i,j} w_{ij} D_{ij} \quad f_i^1, f_i^2$$

(1)

Mô hình biểu diễn ảnh bằng các đặc trưng mức thấp yêu cầu chuẩn hóa các đặc trưng về miền giá trị số $[0,1]$ để bù đắp cho phạm vi khác biệt nhau giữa các thành phần đặc trưng được định nghĩa trong các miền khác nhau.

B. Phép chuẩn hóa các bộ đặc trưng mức thấp

Phép chuẩn hóa giá trị của các đặc trưng mức thấp có ảnh hưởng đến hiệu quả của các hệ thống CBIR [3]. Ngoài các phép chuẩn hóa thông dụng như min-max, 3sigma [2] và gần đây trong [3] các tác giả đã trình bày một phép chuẩn hóa mở rộng của 3sigma để khắc phục giới hạn của yêu cầu dữ liệu phải có phân bố Gauss.

Cho trước một cơ sở dữ liệu (CSDL) đặc trưng mức thấp, sử dụng FCM (về FCM, xem [21, 22, 23]) ta phân cụm CSDL thành C cụm.

Cho $E_{t,i} \quad 1 \leq i \leq n$ và các hằng số $p = p(t) > 1, C = C(t) \in N^+, C \geq 2, m_t = \dim(E_{t,i}), \forall 1 \leq i \leq n$. Thuật toán lặp FCM cực tiểu hóa hàm mục tiêu:

$$J(V, \eta) = \sum_{i=1}^n \sum_{c=1}^C \eta_{c,i}^p \|E_{t,i} - V_c\|^2 \rightarrow \min, \quad (2)$$

với độ đo khoảng cách Euclid, $\|E_{t,i} - V_{t,c}\|^2 = \sum_{j=1}^{m_t} E_{t,i}[j] - V_{t,c}[j]^2$ và các ràng buộc biến như sau:

$$(i) \quad \eta_{c,i} \in [0,1], \forall 1 \leq i \leq n, 1 \leq c \leq C$$

$$(ii) \quad \sum_{c=1}^C \eta_{c,i} = 1, \forall 1 \leq i \leq n \quad (3)$$

$$(iii) \quad \sum_{i=1}^n \eta_{c,i} > 0, \forall 1 \leq c \leq C$$

Phép chuẩn hóa 3 σ -FCM [3] được cho như sau:

$x = x_j \quad m_t, x_{norm} = x_{norm,j} \quad m_t, \forall 1 \leq j \leq m_t$ là vectơ dữ liệu đầu vào theo đặc trưng mức thấp t, vectơ chuẩn hóa x_{norm} của 3 σ -FCM xác định như sau:

$$x_{norm,j} = \frac{\min_{1 \leq c \leq C} \left\{ \frac{x_j - V_{t,c,j}}{3\sigma_{t,c,j}} \right\} + \max_{1 \leq c \leq C} \left\{ \frac{x_j - V_{t,c,j}}{3\sigma_{t,c,j}} \right\}}{C + 1} \quad (4)$$

trong đó tham số C là số cụm của thuật toán phân cụm FCM, độ lệch chuẩn ở cụm c ($1 \leq c \leq C$) trên dữ liệu $E_{t,i} \quad 1 \leq i \leq n$ được tính bằng công thức sau:

$$\forall 1 \leq j \leq m_t, \quad \sigma_{t,c,j} = \sqrt{\frac{\sum_{i=1}^n \eta_{c,i}^p E_{t,i,j} - V_{t,c,j}^2}{\sum_{i=1}^n \eta_{c,i}^p}} = \sqrt{\frac{\sum_{i=1}^n \eta_{t,c,i}^p E_{t,i,j}^2}{\sum_{i=1}^n \eta_{t,c,i}^p} - V_{t,c,j}^2} \quad (5)$$

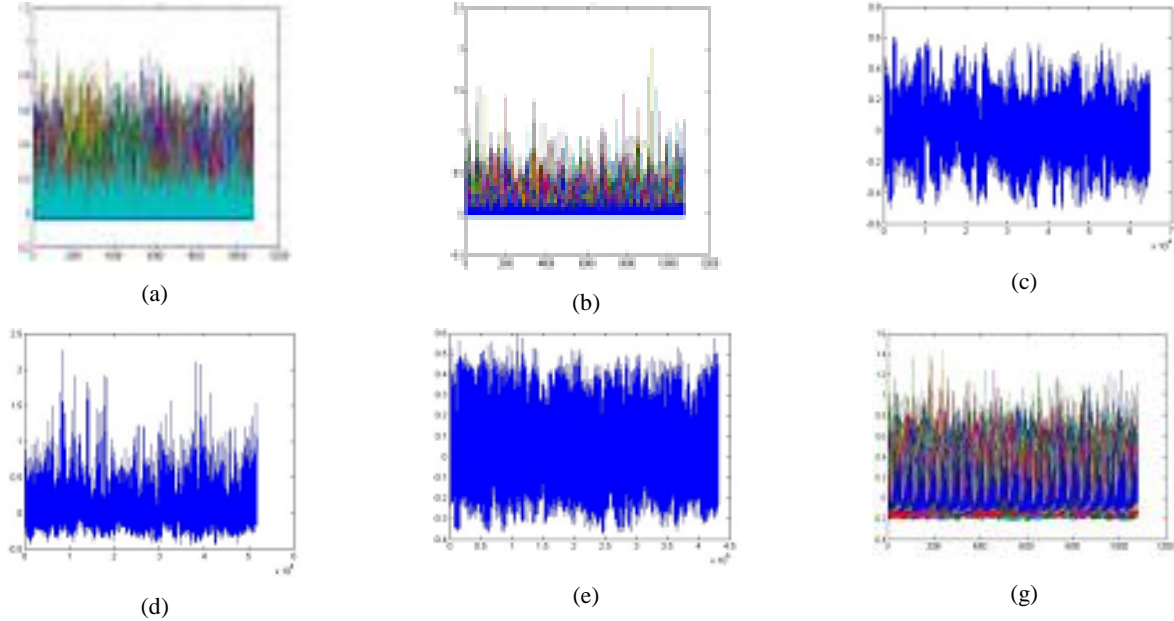
Đặc điểm của 3 σ -FCM :

(1) 3 σ -FCM không cần giả định các thành phần vectơ đầu vào có phân bố Gauss.

(2) 3 σ -FCM bảo toàn thứ tự trên từng thành phần vectơ và hầu hết giá trị thành phần vectơ đã biến đổi nằm trong $[-1,1]$.

(3) Phép chuẩn hóa 3σ-FCM không phải chuẩn hóa từng thành phần vectơ riêng rẽ như phép chuẩn hóa min-max, 3sigma mà thực hiện biến đổi với các thành phần vectơ với tham số biến đổi có liên hệ nhau.

Tuy nhiên 3σ-FCM vẫn còn một số hạn chế, đó là miền giá trị của các thành phần vectơ sau khi biến đổi vẫn có thể hẹp hơn [-1,1] như hình 1a-1g đã chỉ rõ. Ngoài ra, phép biến đổi vẫn chịu ảnh hưởng của tham số số cụm C khi sử dụng thuật toán phân cụm dữ liệu vectơ FCM trước khi biến đổi chuẩn hóa theo công thức (4).



Hình 1. Lược đồ đặc trưng chuẩn hoá theo luật 3σ-FCM. (a) HSV Histogram. (b) autoCorrelogram. (c) GaborTexture. (d) WaveletMoment. (e) Gist [3]

Trong mục III chúng tôi sẽ đề xuất phép biến đổi dựa trên 3σ-FCM vẫn đảm bảo các đặc điểm của phép biến đổi 3σ-FCM nhưng giá trị của các thành phần vectơ sau khi biến đổi trải rộng hơn trong [-1,1] và không phụ thuộc nhiều vào tham số C-số cụm.

III. ĐỀ XUẤT

Sau khi phân cụm FCM dữ liệu vectơ đặc trưng mức thấp t ($t \in \{Color\ moment, LBP, Gabor\ Wavelet\ texture, Edge, Gist\}$ chẳng hạn) của CSDL ảnh ta nhận được tập các vectơ tâm $\{\bar{V}_{t,c}\}_{1 \leq c \leq C(t)}$ và ma trận 2 chiều các giá trị độ thuộc trong đoạn [0,1] là $\{\eta_{t,j,c}\}_{1 \leq j \leq m_t, 1 \leq c \leq C(t)}$, thay vì dùng hệ số tỉ lệ là $K=C(t)+1$ ta sẽ xác định hệ số a, b của biến đổi tuyến tính bảo toàn thứ tự ($a > 0$), $x_j \in [c_j, d_j] \mapsto a_j x_j + b_j \in [-1, 1]$: $a = a_{E_{t,i}} \quad 1 \leq i \leq n \in [0, 1], b = b_{E_{t,i}} \quad 1 \leq i \leq n$ từ độ thuộc từ CSDL ảnh theo tiêu chí cực đại hóa dải tập trung của dữ liệu trong đoạn [-1,1].

Phép chuẩn hóa tối ưu được thiết kế để đảm bảo có không quá $100 * \alpha_{out}$ % vectơ của CSDL đặc trưng mức thấp bộ t tại mỗi thành phần vectơ $j \in \overline{1, \dim(E_t)}$ rơi ra ngoài đoạn [-1,1] và độ trải rộng của tập thành phần j của tập vectơ CSDL trên [-1,1] là lớn nhất.

Định nghĩa 2: Phép chuẩn hóa 3σ-opt với tham số tối ưu.

Giả sử $x = x_j \quad j=1 \dots m_t$ là vectơ dữ liệu đầu vào, $x_{norm} = x_{norm,j} \quad j=1 \dots m_t, \forall 1 \leq j \leq m_t, x_{norm,j}$ là vectơ chuẩn hóa của x,

$$x_{norm,j} \stackrel{def}{=} a_{opt} \left(\min_{1 \leq c \leq C} \left\{ \frac{x_j - V_{t,c,j}}{3\sigma_{t,c,j}} \right\} + \max_{1 \leq c \leq C} \left\{ \frac{x_j - V_{t,c,j}}{3\sigma_{t,c,j}} \right\} \right) + b_{opt}, \tag{6}$$

ở đây a_{opt}, b_{opt} là 2 số thực thỏa mãn:

- (i) $0 < a_{opt} < 1, -0.5 \leq b_{opt} \leq 0.5$
- (ii) a_{opt}, b_{opt} tối ưu hóa hàm mục tiêu sau:

$$F_t(a,b) = \frac{1}{1 + \sqrt{\frac{\sum_{i=1}^n \sum_{1 \leq j \leq \dim(E_t), ad_{t,i,j} + b \in [-1,1]} (ad_{t,i,j} + b)^2}{n * \dim E_t}}} + \# \quad j \in \overline{1, \dim(E_t)} / \# \quad i \in \overline{1, n} / ad_{t,i,j} + b \notin [-1,1] > n * \alpha_{out} \rightarrow \min \quad (7)$$

trong đó

$$d_{t,i,j} \stackrel{def}{=} \min_{1 \leq c \leq C} \left\{ \frac{E_{i,j} - V_{t,c,j}}{3\sigma_{t,c,j}} \right\} + \max_{1 \leq c \leq C} \left\{ \frac{E_{i,j} - V_{t,c,j}}{3\sigma_{t,c,j}} \right\}, t = \overline{1, T}, i = \overline{1, n}, j = \overline{1, \dim(E_t)} \quad (8)$$

$\alpha_{out} \in (0,1)$ là ngưỡng không chế số phần tử thành phần vectơ thứ $j = \overline{1, \dim(E_t)}$ của n vectơ dữ liệu $\{E_{t,i}\}_{i=1}^n$ của CSDL đặc trưng mức thấp t (với dữ liệu có phân bố Gauss chuẩn, phép chuẩn hóa 3σ có $\alpha_{out} = 0.01$).

Nhận xét : Hàm $F_t(a,b)$ được thiết kế theo nguyên tắc tăng thêm giá trị 1.0 tại mỗi thành phần $j \in \overline{1, \dim(E_t)}$ mà có lớn hơn $100 * \alpha_{out} \%$ vectơ của CSDL đặc trưng mức thấp bộ t tại thành phần vectơ rơi ra ngoài đoạn $[-1,1]$. Như vậy $F_t(a,b)$ sẽ tăng khi có nhiều thành phần $j \in \overline{1, \dim(E_t)}$ mà có lớn hơn $100 * \alpha_{out} \%$ vectơ của CSDL đặc trưng mức thấp bộ t tại thành phần vectơ rơi ra ngoài đoạn $[-1,1]$.

Mệnh đề 1:

Nếu $F_t(a,b) < 1$ thì với mỗi thành phần j trong vectơ đặc trưng bộ t , $j = \overline{1, \dim(E_t)}$, số lượng vectơ đặc trưng bộ t trong CSDL $E_{t,i}$ $1 \leq i \leq n$ sau khi biến đổi theo công thức dạng (7), rơi ra ngoài đoạn $[-1,1]$ không vượt quá $n * \alpha_{out}$.

Chứng minh: xem phụ lục.

Thuật toán 1. Phép chuẩn hóa 3σ-opt (Phép chuẩn hóa với tham số tối ưu).

Input: $E_{t,i}$ $1 \leq i \leq n, 1 \leq t \leq T$, hằng số $p = p(t) > 1$, $C = C(t) \in \mathbb{N}^+$, $C \geq 2$, $m_t = \dim(E_{t,i}^F)$, $\forall i = \overline{1, n}$, ngưỡng phần trăm rơi ra ngoài $[-1,1]$ của các thành phần vectơ sau khi chuẩn hóa α_{out} .

Output: $E_{t,i}^{Norm}$ $1 \leq i \leq n$ dữ liệu đã được chuẩn hoá, các tâm $V_{t,c}$ $1 \leq c \leq C_t$, $\sigma_{t,c,j}$ $1 \leq c \leq C_t, 1 \leq j \leq m_t$ và tham số a_{opt}, b_{opt}

Bước 1: FCM C_t, p_t $E_{t,i,c}$ $1 \leq i \leq n, 1 \leq c \leq C_t$ ta được $V_{t,c}$ $c=1$, $\eta_{t,c,i}$ $1 \leq c \leq C_t, 1 \leq i \leq n$, theo công thức (2) và (3)

Bước 2: Tính $\sigma_{t,c,j}$ $1 \leq c \leq C_t, 1 \leq j \leq m_t$ theo công thức (5)

Bước 3: Tính $d_{t,c,j}$ $1 \leq c \leq C_t, 1 \leq j \leq m_t$ theo công thức (8)

Bước 4: Giải tối ưu $F_t(a,b) \rightarrow \min$, xác định $a_{opt} \in (0,1)$, $b_{opt} \in [-0.5,0.5]$ sao cho $F_t(a_{opt}, b_{opt})$ nhỏ nhất, ở đây $F_t(a,b)$ xác định theo công thức (7).

4.1: Khởi đầu $a = \frac{1}{C+1}$, $b = 0$.

4.2: Lặp, thay đổi a và b để $F_t(a,b)$ đạt tới giá trị xấp xỉ giá trị nhỏ nhất (xem [26]).

4.3: Gán $a_{opt} = a$, $b_{opt} = b$.

Bước 5: Chuẩn hóa về $[-1,1]$:

Lặp với mỗi vectơ đặc trưng bộ t của CSDL $\vec{E}_{t,i}$: lặp với mỗi thành phần j $j = \overline{1, m_t}$ của vectơ $\vec{E}_{t,i}$ tính $E_{t,i,j}^{norm}$ theo công thức (6).

Bước 6: Chuẩn hóa về $[0,1]$:

Lặp với mỗi vectơ $\vec{E}_{t,i}^{norm}$: $\forall j = \overline{1, m_t}$ tính lại $E_{t,i,j}^{norm} = \min \left\{ \frac{\max \{E_{t,i,j}^{norm}, -1 + 1\}}{2}, 1 \right\}$.

Trả về: $\vec{E}_{t,i}^{norm}$ $1 \leq i \leq n$, $\vec{V}_{t,c}$ $1 \leq c \leq C_t$, $\sigma_{t,c,j}$ $1 \leq c \leq C_t, 1 \leq j \leq m_t$, a_{opt}, b_{opt} .

Độ phức tạp thuật toán 1 là $O(n)$.

IV. THỰC NGHIỆM

Chúng tôi tiến hành các thí nghiệm trên ba tập dữ liệu Corel10K [24, 27] và Oliva [25]. Đây là các tập dữ liệu chuẩn được sử dụng rộng rãi để đánh giá hiệu năng của các hệ thống CBIR. Các tập dữ liệu này được tổ chức thành các lớp ngữ nghĩa theo cách con người nhận thức về độ tương tự. Mỗi lớp biểu diễn một chủ đề ngữ nghĩa khác nhau, các ảnh trong cùng một lớp được xem là liên quan.

Tập dữ liệu thứ nhất Corel10K, là tập con của cơ sở dữ liệu Corel photo gallery. Nó gồm 10000 ảnh được phân thành 100 lớp với 100 ảnh trên một lớp. Hình 2 chỉ ra một số mẫu ảnh trong tập dữ liệu này.



Hình 2. Một số mẫu trong tập dữ liệu ảnh Corel 10000 ảnh

Tập dữ liệu thứ hai Oliva bao gồm 2600 ảnh được tổ chức thành 8 lớp: Coast & Beach, open country, forest, Mountain, highway street, city center, Tall building như chỉ ra trong hình 3, mỗi lớp có từ 260 đến 409 ảnh.



Hình 3. Một số mẫu trong tập dữ liệu ảnh Oliva

A. Trích chọn đặc trưng

Trong các thí nghiệm, chúng tôi sử dụng 5 đặc trưng toàn cục để mô tả một ảnh: Color Moments, LBP, Gabor Wavelets Texture, Edge, GIST.

Bảng 2. Thuộc tính của các đặc trưng mô tả về màu, kết cấu và hình dạng ảnh

Tên đặc trưng	Đặc điểm biểu diễn	Số chiều vectơ đặc trưng
Color Moments	Màu sắc	81
LBP	Kết cấu	59
Gabor Wavelets Texture	Kết cấu	120
Edge	Hình dạng	37
GIST	Hình dạng	512

Tất cả các đặc trưng này được chuẩn hóa để mỗi thành phần nằm trong khoảng $[-1,1]$ theo thuật toán 1. Khoảng cách Eculid sẽ được sử dụng để tính khoảng cách giữa các đặc trưng.

B. Chỉ số đánh giá

Do phép chuẩn hóa đã đặt ngưỡng số phần tử rơi ra ngoài đoạn $[-1,1]$ của các thành phần vectơ rơi ra ngoài đoạn $[-1,1]$ là không vượt quá α_{out} , một cách tự nhiên, chúng ta có chỉ số đánh giá hiệu quả phép chuẩn hóa vectơ dữ liệu đặc trưng mức thấp bộ t về đoạn $[-1,1]$, cụ thể như sau:

$$RMS \stackrel{def}{=} \sqrt{\frac{\sum_{i=1}^n \sum_{1 \leq j \leq \dim E_i \wedge ad_{i,j} + b \in [-1,1]} (ad_{i,j} + b)^2}{n * \dim E_i}} \quad (9)$$

Chỉ số RMS nói lên độ trải trong đoạn chuẩn hóa $[-1,1]$. RMS càng lớn chứng tỏ phép chuẩn hóa càng tốt theo nghĩa độ trải trong đoạn $[-1,1]$ của mọi tập thành phần $E_{t,i,j}$, $j = 1, \dim(\overline{E}_t)$ là lớn.

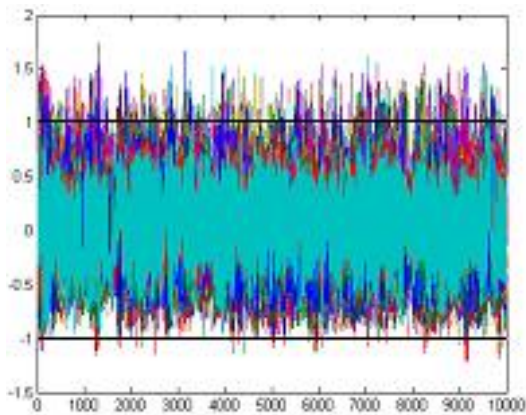
C. Tham số tối ưu tính được và kết quả thực nghiệm chuẩn hóa đề xuất

Bảng 3. Tham số tối ưu của 3σ -opt cho CSDL các đặc trưng mức thấp trong CSDL Corel 10000 mẫu thực hiện với FCM có số cụm $C = 10$, $\alpha_{out}=0.02$ (2%)

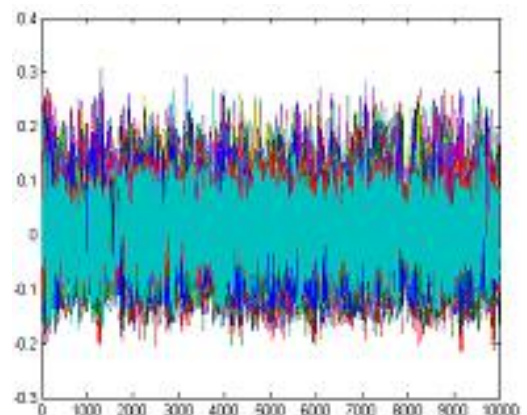
Dữ liệu đặc trưng	\mathbf{a}_{opt}	\mathbf{b}_{opt}
Color Moments	0.5138	0.0096
LBP	0.2315	0.0037
Gabor Wavelets Texture	0.4290	-0.2294
Edge	0.0755	0.500
GIST	0.3377	-0.4953

Bảng 4. Giá trị chỉ số EMR 3σ -opt và 3σ -FCM cho CSDL các đặc trưng mức thấp trong CSDL Corel 10000 mẫu thực hiện với FCM có số cụm $C = 10$

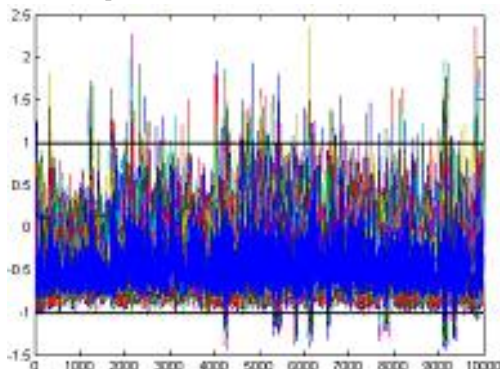
Dữ liệu đặc trưng	RMS 3σ -opt	RMS 3σ -FCM
Color Moments	0.36407	0.065946
LBP	0.28731	0.11644
Gabor Wavelets Texture	0.3881	0.089657
Edge	0.53337	0.12538
GIST	0.54033	0.060937



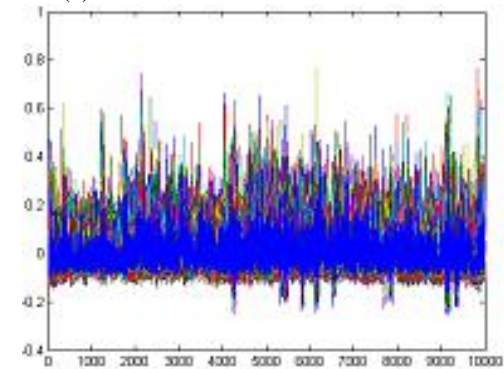
(a) 3σ -opt đề xuất cho color moments 81 chiều



(b) 3σ -FCM cho color moments 81 chiều



(c) 3σ -opt đề xuất cho Gist 512 chiều



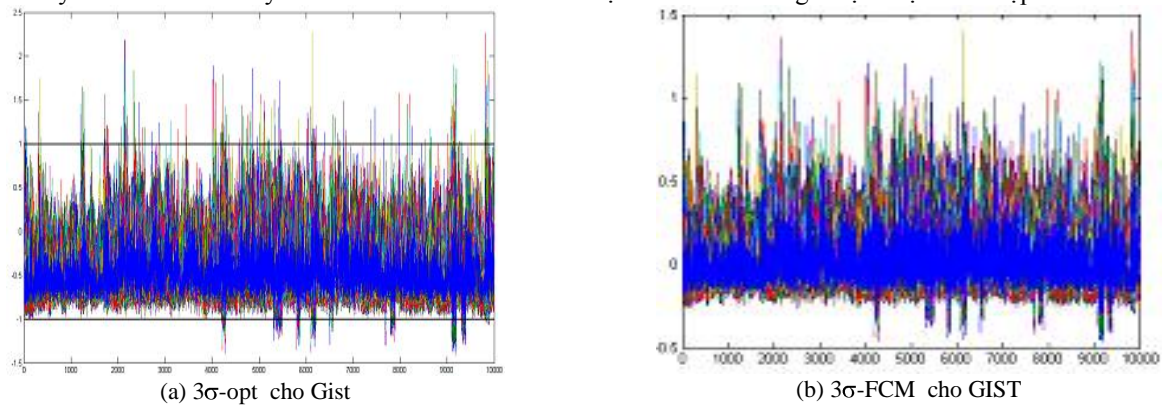
(d) 3σ -FCM cho Gist 512 chiều

Hình 4. Chuẩn hóa cho tập Corel 10000 ảnh, số cụm $C = 10$

3σ -opt có trung bình (a) 50, (c) 9 véctơ thành phần rơi ra ngoài $[-1,1]$, trải rộng trên $[-1,1]$

3σ -FCM có trung bình (b) 0, (c) 0 véctơ thành phần rơi ra ngoài $[-1,1]$, đoạn hẹp của $[-1,1]$

Phép chuẩn hóa 3σ -FCM phụ thuộc mạnh vào tham số số cụm, tuy nhiên 3σ -opt thì không phụ thuộc nhiều vào tham số này. Hình sau cho thấy rõ điều đó khi chuẩn hóa dữ liệu vectơ Gist với giá trị số cụm thiết lập là 5.



Hình 5. Chuẩn hóa cho tập Corel 10000 ảnh, số cụm $C = 5$

So sánh hình 5.c và hình 6.b rõ ràng 3σ -FCM phụ thuộc vào số cụm và khi số cụm là 5 độ trải rộng của 3σ -FCM tốt hơn khi số cụm là 10. Tuy nhiên 3σ -opt thì không phụ thuộc số cụm. Điều này là do bản chất hàm mục tiêu $F_1(a,b)$ ở công thức (7) được tối ưu để trải rộng dữ liệu thành phần vectơ trên đoạn $[-1,1]$.

D. So sánh thực nghiệm trong CBIR

Để so sánh chúng tôi sử dụng phép đánh hạng EMR [20] để thực nghiệm tính hiệu quả của thuật toán chuẩn hóa tối ưu đề xuất của chúng tôi so với [3].



(a) 109020.jpg: Kích thước 768x512



(b) 172023.jpg: Kích thước 768x512



(c) 170012.jpg: Kích thước 512x768

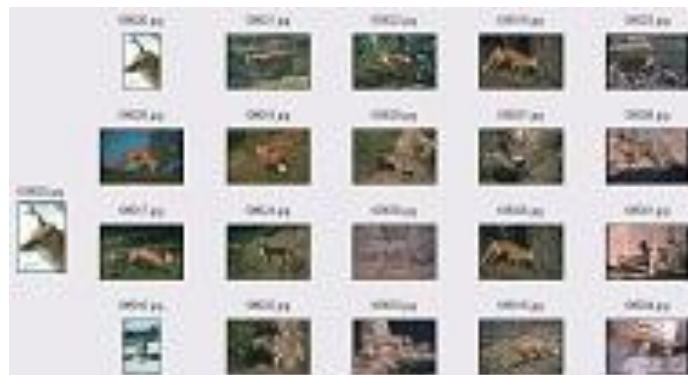


(d) 107018.jpg: Kích thước 512x768

Hình 6. Một số ảnh trong Corel 10000 ảnh



Hình 7. Truy vấn hình ảnh 6.a sử dụng chuẩn hóa 3σ -FCM với EMR, xuất hiện 3 ảnh không liên quan trong 20 ảnh có thứ hạng cao nhất



Hình 8. Truy vấn hình ảnh 6.a sử dụng chuẩn hóa 3σ -opt đề xuất với EMR, tất cả trong số 20 ảnh có thứ hạng cao nhất đều liên quan



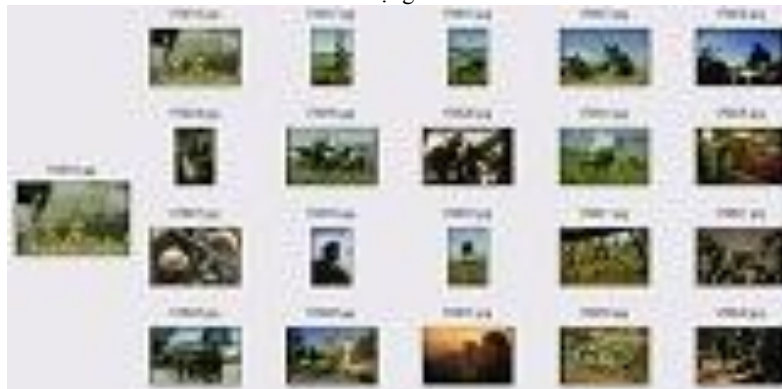
Hình 9. Truy vấn hình ảnh 6.b sử dụng chuẩn hóa 3σ -FCM với EMR, xuất hiện 6 ảnh không liên quan trong 20 ảnh có thứ hạng cao nhất



Hình 10. Truy vấn hình ảnh 6.b sử dụng chuẩn hóa 3σ -opt đề xuất với EMR, tất cả trong số 20 ảnh có thứ hạng cao nhất đều liên quan



Hình 11. Truy vấn hình ảnh 6.c sử dụng chuẩn hóa 3σ -FCM với EMR, xuất hiện 3 ảnh không liên quan trong 20 ảnh có thứ hạng cao nhất



Hình 12. Truy vấn hình ảnh 6.c sử dụng chuẩn hóa 3σ -opt đề xuất với EMR, tất cả trong số 20 ảnh có thứ hạng cao nhất đều liên quan



Hình 13. Truy vấn hình ảnh 6.d sử dụng chuẩn hóa 3σ -FCM với EMR, xuất hiện 3 ảnh không liên quan trong 20 ảnh có thứ hạng cao nhất



Hình 14. Truy vấn hình ảnh 6.d sử dụng chuẩn hóa 3σ -opt đề xuất với EMR, tất cả trong số 20 ảnh có thứ hạng cao nhất đều liên quan

Thực nghiệm với 100 ảnh có thứ hạng cao nhất theo độ đo tương tự EMR trên 10% ảnh truy vấn được chọn ngẫu nhiên từ tập Corel và tập Oliva đều cho kết quả trung bình của 3σ -opt là cao hơn so với 3σ -FCM.

Bảng 6. So sánh hiệu quả truy vấn của 3σ -opt và 3σ -FCM trên 2 tập Corel và Oliva

Tập ảnh	Độ chính xác trung bình	Độ chính xác trung bình
Corel 10000 ảnh	65,3%	69,1%
Oliva	73,3%	75,4%

Các kết quả minh họa thực nghiệm trên cho thấy phép chuẩn hóa 3σ -opt cho hiệu quả hơn 3σ -FCM. Khi các dữ liệu đặc trưng mức thấp được chuẩn hóa với độ khớp cao với đoạn chuẩn hóa $[-1,1]$ độ đo tương tự EMR đã được học ma trận trọng số tối ưu hơn và dẫn đến kết quả đánh hạng được cải thiện.

V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp chuẩn hóa dữ liệu véctơ đặc trưng mức thấp, bảo toàn thứ tự ở các thành phần véctơ sau khi chuẩn hóa dựa trên phép chuẩn hóa 3σ -FCM của các tác giả trong [3], nhưng có ưu điểm về độ trải rộng trên đoạn chuẩn hóa $[-1,1]$, không phụ thuộc tham số C- số cụm được sử dụng cùng với thuật toán FCM.

Ngoài ra bài báo cũng đề xuất một chỉ số đánh giá hiệu quả của một phép chuẩn hóa dữ liệu véctơ m chiều về khối hộp vuông $[-1,1]^m$.

Thực nghiệm với các CSDL ảnh lớn như Corel 10000 mẫu, dựa trên trực quan và chỉ số đánh giá khách quan, kết quả truy vấn ảnh thực tế đã chứng tỏ hiệu quả của phép chuẩn hóa tối ưu đề xuất là vượt trên các thuật toán chuẩn hóa đã có.

TÀI LIỆU THAM KHẢO

- [1] Cong, Kidiyo Kpalma, and Joseph Ronsin. "Color textured image retrieval by combining texture and color features." Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European. IEEE, 2012.
- [2] Rui, Yong, et al. "Relevance feedback: a power tool for interactive content-based image retrieval." Circuits and Systems for Video Technology, IEEE Transactions on 8.5 (1998): 644-655.
- [3] Vũ Văn Hiệu, Ngô Hoàng Huy, Ngô Quốc Tạo, Nguyễn Hữu Quỳnh, Một phương pháp mới chuẩn hóa dữ liệu và hiệu chỉnh trọng số cho tổ hợp đặc trưng trong tra cứu ảnh theo nội dung. Tạp chí Công nghệ Thông tin và Truyền thông, tập V-1, Số 15(35) 6-2016, trang 63-75.
- [4] Ciocca, Gianluigi, and Raimondo Schettini. "A relevance feedback mechanism for content-based image retrieval." Information processing & management 35.5 (1999): 605-632.
- [5] Mehrotra, Sharad, et al. "Multimedia analysis and retrieval system." Proc. of The 3rd Int. Workshop on Information Retrieval Systems. 1997.
- [6] Ortega, Michael, et al. "Supporting similarity queries in MARS." Proceedings of the fifth ACM international conference on Multimedia. ACM, 1997.
- [7] Ngo Truong Giang, Ngo Quoc Tao, Nguyen Duc Dung and Ngo Hoang Huy, LEARNING INTERACTION MEASURE WITH RELEVANCE FEEDBACK IN IMAGE RETRIEVAL. Journal of Computer Science and Cybernetics.
- [8] Smith, John R., and Shih-Fu Chang. "VisualSEEK: a fully automated content-based image query system." Proceedings of the fourth ACM international conference on Multimedia. ACM, 1997.
- [9] Gudivada, Venkat N., and Vijay V. Raghavan. "Content based image retrieval systems." Computer 28.9 (1995): 18-22.
- [10] Deng, Yining, et al. "An efficient color representation for image retrieval." Image Processing, IEEE Transactions on 10.1 (2001): 140-147.
- [11] Jose, Sebin, and Philumon Joseph. "Content based Image Retrieval System with Watermarks and Relevance Feedback." International Journal of Computer Applications 99.11 (2014): 1-6.
- [12] Swain, Michael J., and Dana H. Ballard. "Color indexing." International journal of computer vision 7.1 (1991): 11-32.
- [13] Rui, Yong, Thomas S. Huang, and Sharad Mehrotra. "Content-based image retrieval with relevance feedback in MARS." Image Processing, 1997. Proceedings., International Conference on. Vol. 2. IEEE, 1997.
- [14] Tamura, Hideyuki, Shunji Mori, and Takashi Yamawaki. "Textural features corresponding to visual perception." Systems, Man and Cybernetics, IEEE Transactions on 8.6 (1978): 460-473.
- [15] Mao, Jianchang, and Anil K. Jain. "Texture classification and segmentation using multiresolution

- [16] Ohanian, Philippe P., and Richard C. Dubes. "Performance evaluation for four classes of textural features." Pattern recognition 25.8 (1992): 819-833.
- [17] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. 29(1):51–59, January 1996.
- [18] Jianke Zhu, Steven C.H. Hoi, Michael R. Lyu and Shuicheng Yan, "Near-Duplicate Keyframe Retrieval by Nonrigid Image Matching," ACM Multimedia' 2008.
- [19] L. Fei-Fei, R. Fergus and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004.
- [20] Bin Xu, Jiajun Bu, Chun Chen, Can Wang, Deng Cai, and Xiaofei He, "EMR: A Scalable Graph-based Ranking Model for Content-based Image Retrieval", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 1, JANUARY 2015.
- [21] Bezdek, James C. Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, 2013.
- [22] Bhanu, Bir, and Anlei Dong. "Concepts learning with fuzzy clustering and relevance feedback." Engineering Applications of Artificial Intelligence 15.2 (2002): 123-138.
- [23] Yang, Miin-Shen, Pei-Yuan Hwang, and De-Hua Chen. "Fuzzy clustering algorithms for mixed feature variables." Fuzzy Sets and Systems 141.2 (2004): 301-317.
- [24] Z.-H. Zhou, K.-J. Chen, and H.-B. Dai. Enhancing relevance feedback in image retrieval using unlabeled data. ACM Transactions on Information Systems, 24(2):219–244, 2006.
- [25] Hoi, S.C.H.; Lyu, M.R.; Rong Jin, "A unified log-based relevance feedback scheme for image retrieval," in Knowledge and Data Engineering, IEEE Transactions on , vol.18, no.4, pp.509-524, April 2006.
- [26] <https://www.mathworks.com/help/matlab/math/optimizing-nonlinear-functions.html>
- [27] <http://corel.digitalriver.com>

PHỤ LỤC A. Một số phép chuẩn hoá hay được sử dụng trong CBIR [3]

Chuẩn hóa min-max về [0,1]:

$$f_i = f_{i,j} \mapsto f'_i = f'_{i,j}, \forall j = \overline{1, \dim(f_i)}, f'_{i,j} = \frac{f_{i,j} - \min_{E_i} E_{i,j}}{\max_{E_i} E_{i,j} - \min_{E_i} E_{i,j}} \quad (\text{A.1})$$

Chuẩn hóa 3σ về [0,1] (hầu hết):

$$f_i = f_{i,j} \mapsto f'_i = f'_{i,j}, \forall j = \overline{1, \dim(f_i)}, f'_{i,j} = 0.5 * \left\{ \frac{f_{i,j} - m_j}{3\sigma_j} + 1 \right\}, \quad (\text{A.2})$$

Trong đó $m_j \stackrel{\text{def}}{=} \text{mean } E_{i,j}$, $\sigma_j \stackrel{\text{def}}{=} \text{var } E_{i,j}$, ở đây $\overline{E_i}$ là một CSDL đặc trưng mức thấp của các ảnh.

Chuẩn hóa theo min – max làm cho hầu hết thông tin hữu ích bị chuyển vào một phạm vi rất hẹp trong [0,1] nếu giá trị max lớn, 3σ rải đều trong [-1,1] nhưng yêu cầu dữ liệu là một chuỗi Gaussian.

Tính chất : Chuẩn hóa theo min-max, 3σ bảo toàn thứ tự $f_{1,i}[j] \geq f_{2,i}[j] \Rightarrow f'_{1,i}[j] \geq f'_{2,i}[j]$

Luật 3σ : x là $N(\mu, \sigma)$, thì xác suất $P\left(\left|\frac{x-\mu}{\sigma}\right| \leq 1\right) \geq 0.99$ [2]

Hạn chế: Chuẩn hóa theo min – max làm cho hầu hết thông tin hữu ích bị chuyển vào một phạm vi rất hẹp trong [0,1] nếu giá trị max lớn, 3σ rải đều trong [-1,1] nhưng yêu cầu dữ liệu là một chuỗi Gaussian.

PHỤ LỤC B

Mệnh đề 1:

Nếu $F_t(a,b) < 1$ thì với mỗi thành phần j trong véctor đặc trưng bộ t , $j = \overline{1, \dim(E_t)}$, số lượng véctor đặc trưng bộ t trong CSDL $E_{t,i}$ $1 \leq i \leq n$ sau khi biến đổi theo công thức dạng (7), rơi ra ngoài đoạn $[-1,1]$ không vượt quá $n^* \alpha_{out}$.

Chứng minh:

$F_t(a,b) < 1$ nên $\# j \in \overline{1, \dim(E_t)} / \# i \in \overline{1, n} / ad_{t,i,j} + b \notin [-1,1] > n^* \alpha_{out} = 0$,

nghĩa là tập $j \in \overline{1, \dim(E_t)} / \# i \in \overline{1, n} / ad_{t,i,j} + b \notin [-1,1] > n^* \alpha_{out} = \emptyset$.

Nên $\forall j \in \overline{1, \dim(E_t)}$, ta có $\# i \in \overline{1, n} / ad_{t,i,j} + b \notin [-1,1] \leq n^* \alpha_{out}$. Tức là số véctor thứ i , $d_{t,i,j}$ mà $ad_{t,i,j} + b \notin [-1,1]$ không vượt quá $n^* \alpha_{out}$.

A NOVEL LOW LEVEL FEATURE NORMALISATION METHOD FOR CONTENT BASED IMAGE RETRIEVAL

Hoang Xuan Trung, Doan Van Hoa, Nguyen Tan An, Ngo Hoang Huy

ABSTRACT: In CBIR, the image is represented by many low-level features that describe the color, texture and shape of the image. Image query algorithms often construct global distance measurements between feature véctos representing an image object. The combination of different image features in a global distance measurement requires standardized data sets. The standardization of image-specific data has affected CBIR quality, and there have been several standardization uses such as min-max normalization, 3-sigma standardization, and 3 σ -FCM extended standardization. This presents a new standardized method of upgrading from 3 σ -FCM standardization. Experimentation has shown the effectiveness of the proposed algorithm for image normalization of image data. The dynamic range of the low level features data is normalized on the $[0,1]$ segment that is wider than the corresponding 3sigma-FCM normalization interval. The experiment also demonstrates that the new standardization also increases CBIR quality when combined with algorithms for measuring analogue images such as EMR.

Keywords: CBIR, Low level features, FCM, 3sigma-FCM, 3 σ -opt, EMR.