MỘT PHƯƠNG PHÁP PHÂN LỚP CHO BÀI TOÁN TÌM KIẾM ẢNH DỰA TRÊN THUẬT TOÁN k-NN

Huỳnh Thị Châu Lan*, Lê Hữu Hà, Nguyễn Hải Yến

Trường Đại học Công nghiệp Thực phẩm TP.HCM

*Email: *lanhtc@hufi.edu.vn*

Ngày nhận bài: 06/7/2020; Ngày chấp nhận đăng: 27/8/2020

TÓM TẮT

Trong bài báo này, một tiếp cận phân lớp dữ liệu được thực hiện nhằm áp dụng cho bài toán tìm kiếm ảnh tương tự qua đặc trưng thị giác túi từ BoVW (Bag of Visual Words). Phương pháp phân lớp được thực hiện dựa trên thuật toán k-NN (k-Nearest Neighbor) với dữ liệu đầu vào là một véc-tơ đặc trưng của hình ảnh. Từ một tập dữ liệu ảnh ban đầu, chúng tôi xây dựng một cấu trúc túi từ thị giác lưu trữ các hình ảnh có đặc trưng tương đồng theo nội dung. Dựa trên việc phân lớp một hình ảnh đầu vào theo phương pháp k-NN, một tập các hình ảnh được trích xuất từ cấu trúc túi từ thị giác. Trong phương pháp k-NN, ngoài k phần tử láng giềng gần nhất thì một bán kính được sử dụng để thống kê các phân lớp của hình ảnh. Mỗi một túi từ chứa nhiều hình ảnh tương đồng về nội dung và có nhiều phân lớp ngữ nghĩa khác nhau; đồng thời, mỗi túi từ liên kết đến các túi từ khác qua phân lớp ngữ nghĩa đại diện. Thực nghiệm được xây dựng trên bộ ảnh COREL (1.000 ảnh) nhằm đánh giá độ chính xác đồng thời so sánh với các công trình khác trên cùng bộ dữ liệu. Theo kết quả thực nghiệm, những đề xuất của nhóm tác giả là hiệu quả và có thể áp dụng trong các hệ thống đa phương tiện khác nhau.

Từ khóa: k-NN, phân lớp, túi từ, ảnh tương tự, độ đo tương tự.

1. GIỚI THIỆU

Theo số liệu thống kê của tập đoàn dữ liệu quốc tế IDC (*International Data Corporation*), năm 2018 dung lượng dữ liệu toàn cầu khoảng 33 zettabyte (1 zettabyte = 1 nghìn tỷ gigabyte), ước tính đến năm 2025 có khoảng 175 zettabyte; trong đó, 90 zettabyte được tạo ra từ các thiết bị IoT, 49% dữ liệu được lưu trữ trên môi trường đám mây, gần 30% dữ liệu sẽ được sử dụng để xử lý theo thời gian thực [1, 2].

Mặt khác, dữ liệu đa phương tiện (văn bản, hình ảnh, âm thanh và video) đã được phát triển nhanh chóng trên nhiều hệ thống khác nhau, như: điện thoại thông minh, hệ thống mô phỏng đối tượng 2D, 3D, WWW, và các thiết bị viễn thông... Năm 2015, tổng số hình ảnh toàn cầu đạt 3,2 nghìn tỷ; năm 2016, có 3,5 triệu hình ảnh được chia sẻ trong mỗi phút và có 2,5 nghìn tỷ hình ảnh được chia sẻ và lưu trữ trực tuyến. Trong năm 2017, thế giới đã tạo ra 1,2 nghìn tỷ hình ảnh và tổng số ảnh toàn cầu đến năm 2017 là 4,7 nghìn tỷ; trong đó, các hình ảnh được tạo ra từ thiết bị mobile là 90% [3]. Ảnh số đã trở nên thân thuộc với cuộc sống của con người và được ứng dụng trong nhiều hệ thống tra cứu thông tin đa phương tiện như Hệ thống thông tin bệnh viện (Hospital Information System), Hệ thống thông tin địa lý (Geographic Information System), Hệ thống thư viện số (Digital Library System), ứng dụng y sinh, trong giáo dục đào tạo, giải trí… [4, 5].

Kích thước cũng như số lượng ảnh ngày càng tăng nên cần phải có các hệ thống truy vấn ảnh trên các thiết bị cũng như trong các hệ thống đa phương tiện. Việc tra cứu ảnh để tìm ra

tập ảnh tương tự và phân loại hình ảnh là một trong những bài toán quan trọng của nhiều hệ thống đa phương tiện [6].

Việc tra cứu ảnh có nhiều giai đoạn chính, bao gồm: tiền xử lý ảnh, rút trích đặc trưng, phân cụm dữ liệu hình ảnh, phân lớp đối tượng, tìm kiếm tập ảnh tương tự [7, 8]. Trong cách tiếp cận của nhóm tác giả, kỹ thuật phân lớp *k*-NN được áp dụng cho bài toán tìm kiếm ảnh dựa trên kỹ thuật chọn phần tử láng giềng và các túi từ thị giác BoVW (Bag of Visual Word) nhằm giảm chi phí tính toán và tăng tốc độ tìm kiếm hình ảnh.

Đóng góp của bài báo là: (1) cải tiến thuật toán k-NN nhằm phân lớp dữ liệu để tạo ra các phân loại ngữ nghĩa cho hình ảnh, (2) xây dựng cấu trúc túi từ thị giác để tìm kiếm hình ảnh tương tự, (3) thiết kế mô hình tìm kiếm ảnh tương tự dựa trên việc kết hợp thuật toán k-NN và túi từ thị giác BoVW, (4) xây dựng thực nghiệm và minh chứng tính đúng đắn của đề xuất trên một bộ dữ liệu ảnh thông dụng.

Phần còn lại của bài báo gồm: Phần 2 khảo sát và phân tích ưu nhược điểm của các công trình liên quan để chứng minh tính khả thi của bài toán phân lớp và tìm kiếm ảnh tương tự; Phần 3 trình bày thuật toán phân lớp k-NN và phương pháp tìm kiếm ảnh tương tự dựa trên túi từ thị giác; Thực nghiệm được mô tả trong phần 4 và kết quả được đánh giá trên bộ dữ liệu ảnh COREL (1.000 ảnh); Phần 5 là kết luận và hướng phát triển tiếp theo.

2. CÁC CÔNG TRÌNH LIÊN QUAN

Gần đây, nhiều công trình sử dụng phương pháp phân lớp dựa trên kỹ thuật *k*-NN nhằm thực hiện bài toán phân lớp và tìm kiếm ảnh như: Truy xuất hình ảnh dựa trên nội dung cho bài toán nhận dạng nhiều đối tượng trái cây bằng cách sử dụng k-Means và *k*-NN [9]; Phương pháp trích xuất đặc trưng SIFT để mô tả đặc trưng hình ảnh và được áp dụng trong hệ CBIR kết hợp phân lớp trên mạng BayesNet và *k*-NN [10]; Một phương pháp học có giám sát để tạo chỉ mục cho hình ảnh dựa trên phương pháp xấp xỉ láng giềng gần nhất bằng *k*-NN [11]; Một cách tiếp cận khác sử dụng *k*-NN kết hợp với trọng số nhằm thực hiện chú thích hình ảnh tự động [12]; Một phương pháp chọn lựa đặc trưng sử dụng kỹ thuật học có giám sát *k*-NN trong hệ thống CBIR [13]; Kết hợp thuật toán K-Means và *k*-NN để phân loại ảnh về trái cây [14].

Năm 2014, Xiaohui và cộng sự đã xây dựng độ đo tương tự dựa trên ràng buộc không gian giữa các đối tượng đặc trưng để từ đó thực hiện bài toán tìm kiếm ảnh. Trong phương pháp này, nhóm tác giả thực hiện việc kết hợp giữa phương pháp *k*-NN và túi từ thị giác để truy vấn ảnh. Trong túi từ thị giác, các hình ảnh được thống kê và gom nhóm theo kỹ thuật phân lớp *k*-NN để tạo ra nhóm các hình ảnh tương tự nhau. Trong bài báo này, các túi từ thị giác chứa đựng các hình ảnh dựa trên việc phân lớp *k*-NN trong CSDL ban đầu chưa xây dựng được trọng số của mỗi túi từ theo phân lớp các hình ảnh. Hơn nữa, thuật toán *k*-NN được thực hiện trên độ đo của đối tượng đặc trưng và chưa giải quyết việc phân lớp hình ảnh trong trường hợp số lượng các phân lớp cân bằng nhau [15].

Imran và cộng sự (2014) đã đề xuất hệ thống CBIR mới bằng cách kết hợp 2 đặc trưng màu sắc và kết cấu. Trong đó, bố cục màu (CLD) từ MPEG-7 được sử dụng để trích xuất màu và độ đo trung bình, phương sai, độ lệch và entropy được sử dụng làm bộ mô tả kết cấu. Kết quả thực nghiệm trên bộ ảnh COREL và được so sánh với 4 hệ thống uy tín khác (SIMPLIcity, dựa trên biểu đồ, FIRM và Phân đoạn phương sai) để đánh giá hiệu năng của phương pháp do nhóm tác giả đề xuất [16].

Li và Mooi (2015) đã xây dựng túi từ thị giác dựa trên lược đồ màu sắc và chọn những hình ảnh đưa vào túi từ dựa trên màu sắc của số lượng điểm ảnh. Với mỗi hình ảnh đầu vào được phân loại dựa trên túi từ thị giác này và lấy các hình ảnh lân cận của các ảnh gần nhất trong túi từ để truy xuất tập ảnh tương tự trong CSDL ảnh ban đầu. Việc truy xuất tập ảnh tương tự được thực hiện bằng phương pháp *k*-NN. Trong phương pháp này, nhóm tác giả thực

hiện hai pha của phương pháp *k*-NN kết hợp với túi từ thị giác nhưng vẫn chưa xây dựng được mối quan hệ giữa các túi từ [17].

Huneiti và cộng sự (2015) đề xuất một phương pháp CBIR bằng cách trích xuất các véc-tơ đặc trưng màu và kết cấu, sử dụng phép biến đổi Wavelet rời rạc (Discrete Wavelet Transform) và mạng (SOM). Các hình ảnh được phân nhóm theo màu sắc, với mỗi hình ảnh truy vấn, các véc-tơ đặc trưng kết cấu được so sánh dựa vào độ đo tương tự Euclide để truy xuất tập các hình ảnh tương tự. Ngoài ra, các hình ảnh có liên quan khác cũng được truy xuất bằng cách sử dụng vùng lân cận của hình ảnh tương tự nhất từ tập dữ liệu được phân nhóm thông qua mạng SOM. Thực nghiệm được thực hiện trên bộ ảnh COREL, nhưng hiệu suất chưa cao do việc phân loại màu sắc từ đầu mà không thực hiện so sánh véc-tơ đặc trưng màu sắc [18].

Shrinivasacharya và cộng sự (2015) đề xuất một kỹ thuật trích xuất đặc trưng sử dụng cách tiếp cận kết hợp kỹ thuật dò cạnh và kỹ thuật lọc trung vị để trích xuất các đặc điểm từ hình ảnh. Bên cạnh đó, nhóm tác giả sử dụng kỹ thuật SOM để phân cụm các đối tượng ảnh đã trích xuất đặc trưng. Trên cơ sở đó, một hệ thống truy vấn ảnh được xây dựng dựa trên bản đồ tự tổ chức và trả về tập ảnh tương tự với ảnh truy vấn. Thực nghiệm được đánh giá trên bộ ảnh Corel-1000 [19].

Erwin và cộng sự (2017) đề xuất hệ thống nhận dạng trái cây được xử lý qua 3 bước: đầu tiên là trích xuất các đặc trưng, sau đó thực hiện gom cụm bằng phương pháp K-Means và cuối cùng sử dụng kỹ thuật k-NN để phân lớp. Theo kết quả thực nghiệm, hệ thống phân lớp đạt được độ chính xác 92,5% cho ảnh đơn đối tượng, 90% cho ảnh đa đối tượng [9]. Tuy nhiên, hệ thống chỉ nhận diện trên các bộ ảnh về trái cây, thuật toán K-Means được áp dụng theo phương pháp centroid và phải cập nhật tâm cụm khi dữ liệu thay đổi, chưa xử lý trường hợp số lượng láng giềng có số phân lớp bằng nhau.

Zhang và cộng sự (2017) đề xuất một thuật toán xếp hạng các ảnh đa nhãn dựa trên mô hình *k*-NN. Thuật toán dựa vào xác suất của nhãn kết hợp với các mẫu lân cận xung quanh mẫu truy vấn. Trong cách tiếp cận này, các mẫu tích cực được xem xét và xếp hạng. Nhóm tác giả đã sử dụng bốn bộ ảnh đa nhãn phổ biến để đánh giá thuật toán đề xuất và kết quả cho thấy hiệu suất đạt được tốt hơn so với các phương pháp khác [14]. Trong phương pháp này, nhóm tác giả chỉ áp dụng sắp xếp theo một ảnh đầu vào cho trước và không tạo ra được một cấu trúc để tìm một tập các hình ảnh tương tự.

Kumar và cộng sự (2018) sử dụng phương pháp trích xuất đặc trưng ảnh SIFT (Scale Invariant Feature Transform). Trong đó, SIFT là phép trích xuất đặc trưng đối tượng và bất biến đối với phép biến đổi theo tỷ lệ, quay… Từ đó, đặc trưng này được sử dụng để tìm kiếm ảnh theo nội dung dựa trên phương pháp k-NN. Kết quả thực nghiệm của hệ thống đạt được độ chính xác 88,9% trên bộ ảnh Wang [10]. Tuy nhiên, hệ thống chưa thực hiện được việc phân lớp nếu như số lượng các láng giềng thuộc mỗi lớp xấp xỉ nhau.

Shichao và cộng sự (2019) đề xuất phương pháp học có giám sát để đánh chỉ mục cho ảnh dùng *k*-NN, một thuật toán gán lại nhãn cho các ảnh huấn luyện được đề xuất nhằm thiết lập mối quan hệ giữa các loại nhãn ảnh và các mã từ. Từ đó, một bộ dữ liệu huấn luyện để phân lớp nhằm mở rộng tập các mẫu. Thực nghiệm cho thấy hệ thống dùng phương pháp học có giám sát để tạo chỉ mục cho kết quả tốt hơn mô hình sử dụng phương pháp học không giám sát trên cùng bộ dữ liệu thử nghiệm (MNIST, CIFAR-10) [11]. Tuy nhiên, phương pháp này có 2 hạn chế: Một là, khi gán nhãn đối tượng có thể bị nhầm lẫn vì sử dụng phương pháp *k*-NN để chọn láng giềng gần nhất nhằm tạo chỉ mục cho hình ảnh; hai là, chỉ sử dụng độ đo tương tự làm tiêu chuẩn cho quá trình đối sánh, điều này dẫn đến việc gán mã từ sai cho một đặc trưng hình ảnh nhất định.

Yanchun và cộng sự (2019) đưa ra mô hình *k*-NN có trọng số (*weight k-NN*) kết hợp phương pháp phân biệt tuyến tính đa nhãn để phân lớp đối tượng dựa trên trọng số nhằm cải

thiện độ chính xác trong việc tính toán dự đoán ngữ nghĩa đối tượng hình ảnh [12]. Qua thực nghiệm của hệ thống cho thấy, hệ thống đã thực thi hiệu quả trên các tập dữ liệu lớn. Tuy nhiên, phương pháp này tốn thời gian trong pha huấn luyện và gán nhãn lớp cho hình ảnh, vẫn chưa xây dựng một cấu trúc tìm kiếm ảnh tương tự theo nội dung để tăng tính hiệu quả về thời gian.

Alqasemi và các cộng sự (2019) đề xuất một hướng tiếp cận tìm kiếm ảnh theo nội dung dựa trên kỹ thuật *k*-NN kết hợp với các đặc trưng thống kê trên mỗi hình ảnh trong không gian RGB và đánh giá độ tương tự dựa trên độ đo Euclide. Trong pha tìm kiếm ảnh tương tự được thực hiện bằng cách lấy các nhóm ảnh tương tự với ảnh truy vấn [13]. Đề xuất của nhóm mang lại tính hiệu quả và đơn giản hóa hệ thống tìm kiếm ảnh theo nội dung. Trong bài báo này, việc tìm kiếm các nhóm ảnh tương tự thực hiện một cách tuyến tính và không có cấu trúc tìm kiếm các nhóm ảnh láng giềng để mở rộng cho bài toán tìm kiếm ảnh tương tự.

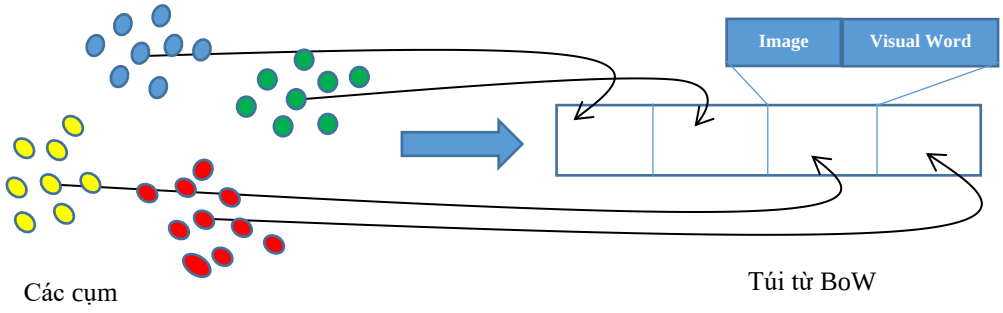
Shuang và cộng sự (2020) kết hợp thuật toán gom cụm K-Means và túi từ thị giác để tìm kiếm tập ảnh tương tự, trong đó túi từ thị giác được xây dựng dựa trên việc gom nhóm các đặc trưng theo thị giác để hình thành các túi từ lưu trữ các từ thị giác của hình ảnh. Ứng với mỗi hình đưa vào được trích xuất đặc trưng và tìm độ tương tự với các túi từ gần nhất để trích xuất ra tập ảnh tương tự [20]. Nhóm tác giả đã sử dụng thuật toán K-Means và túi từ thị giác để tìm kiếm ảnh tương tự, đồng thời đưa ra các ngữ nghĩa tương ứng với túi từ. Trong phương pháp này, các nhóm túi từ là độc lập và chưa phân lớp được nội dung của mỗi hình ảnh.

Theo các công trình đã khảo sát như trên, phương pháp tìm kiếm ảnh tương tự theo nội dung dựa trên kỹ thuật BoVW và *k*-NN là hoàn toàn khả thi. Tuy nhiên, các kỹ thuật đã khảo sát vẫn chưa kết hợp và cải tiến giữa 2 cấu trúc này để giải quyết bài toán tìm kiếm ảnh tương tự. Trong bài báo này, nhóm tác giả đề xuất một tiếp cận mới dựa trên mô hình túi từ thị giác kết hợp với kỹ thuật *k*-NN để phân lớp và tìm kiếm một tập ảnh tương tự. Trong mô hình túi từ, các đặc trưng hình ảnh được lưu trữ cùng với phân lớp của hình ảnh và liên kết với các túi từ khác dựa trên trọng số tỷ lệ giữa các phân lớp ưu thế. Sau đó, với mỗi hình ảnh đầu vào được phân lớp bằng kỹ thuật *k*-NN dựa trên *k* láng giềng gần nhất và bán kính cho trước.

3. PHƯƠNG PHÁP TRA CỨU ẢNH

3.1. Túi từ thị giác

Trong bài báo này, nhóm tác giả xây dựng một mô hình túi từ thị giác BoVW có thể phân loại và tìm kiếm ảnh tương tự dựa trên ngữ nghĩa của mỗi hình ảnh trong túi từ. Mỗi túi từ có một từ thị giác đại diện cho nhóm hình ảnh tương tự và giá trị trọng số được lưu trữ để tìm kiếm các túi từ lân cận theo ngữ nghĩa thị giác. Để xây dựng túi từ, thuật toán K-Means được thực hiện để phân cụm tất cả các véc-tơ đặc trưng trong tập dữ liệu ảnh và xác định được giá trị tâm cũng như ngữ nghĩa thị giác của túi từ đó dựa trên tập dữ liệu huấn luyện. Trong Hình 1, phương pháp tạo túi từ tự động từ một CSDL hình ảnh được thực hiện dựa trên việc phân cụm K-Means theo đặc trưng hình ảnh.



Hình 1. Mô tả cách tạo túi từ

Thuật toán CBVW

Đầu vào: Tập dữ liệu ảnh L= {<f_i,v_i> | với f_i, v_i lần lượt là véc-tơ đặc trưng và phân lớp ngữ nghĩa}

Đầu ra: Tập các túi từ được gán nhãn và có trọng số.

Begin

Khởi tạo số túi từ k;

For j = 1 to k do

$\Omega_j = \emptyset$;

$\Omega_j.Center = f_i$;

EndFor

Foreach (<f_i,v_i> $\in L$) do

D_0 = Min {Euclide (f_i , $\Omega_j.Center$), j = 1..k };

$\Omega_j = \Omega_j \cup \{f_i\}$;

Update ($\Omega_i.Center$);

EndForeach

Return Ω ;

End

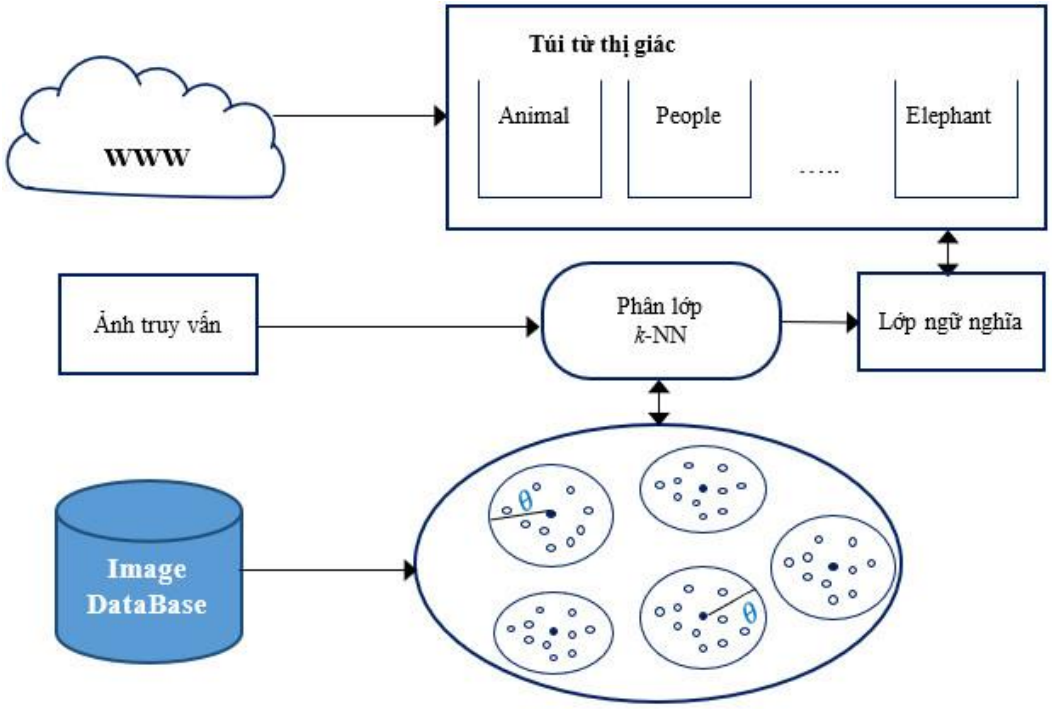
Mệnh đề 1. Độ phức tạp của thuật toán **CBVW** là $O(n * k)$. Với n là số phần tử trong tập véc-tơ đặc trưng trong tập dữ liệu ảnh L, k là số túi từ.

Chứng minh: Ứng với mỗi phần tử f_i trong tập dữ liệu L hệ thống tiến hành tính khoảng cách Euclide giữa nó với k tâm của từng túi từ để tìm ra túi từ phù hợp nhất mà chúng thuộc vào. Do đó, độ phức tạp của thuật toán là $O(n * k)$. ■

Trong thuật toán CBVW, phương pháp gom cụm K-Means được ứng dụng dựa trên các tâm đã chọn. Các phần tử trong tập dữ liệu lần lượt được phân phối vào các túi từ. Dựa trên tập túi từ này, tập ảnh tương tự được trích xuất thông qua phân lớp ngữ nghĩa bằng thuật toán k-NN.

3.2. Thuật toán KNN

Để phân lớp một ảnh đầu vào bằng thuật toán k-NN, một véc-tơ đặc trưng được trích xuất và tìm kiếm các láng giềng gần nhất dựa trên một bán kính đồng thời thống kê theo các phân lớp của k láng giềng gần nhất. Sau khi phân lớp hình ảnh đầu vào, tập hình ảnh tương tự được trích xuất từ các túi từ thị giác.



Hình 2. Mô tả thuật toán *k*-NN kết hợp BoVW

Đầu tiên tập dữ liệu đầu vào sẽ được gom thành *k* cụm theo thuật toán K-Means và *k* véc-tơ tâm tương ứng sẽ làm cơ sở phân lớp cho thuật toán *k*-NN. Tiếp theo, chúng tôi tiến hành xây dựng túi từ dựa trên bộ dữ liệu ảnh ban đầu để thực hiện tìm kiếm tập ảnh tương tự và ngữ nghĩa của ảnh truy vấn. Việc tìm kiếm ảnh tương tự này được thực hiện bằng cách ánh xạ vào từ mã tương ứng trong túi từ.

Thuật toán CkNN

Đầu vào: Một ảnh ***I***, tập đặc trưng ảnh ***F*** đã được gom thành *m* cụm C = {<F_i, I_i> | *i = 1..m*}, bán kính θ

Đầu ra: Lớp ngữ nghĩa ***S*** của ảnh ***I***

Begin

K = ∅ ;

f_I = ExtractFeature(***I***);

dmin = Min {*Euclide* (f_i, I_i), *i = 1..m}.*

If (∃!*dmin*) **then**

S = Classification (f_I, F_i);

Else

Foreach ($f_i \in F_j$) **do** //F_j là các cụm có khoảng cách từ tâm đến *I* là nhỏ nhất

If (Euclide(f_i, f_I) < θ) **then**

K = K ∪ {f_i};

EndIf

EndForeach

S = Classification (f_I, K);

EndIf

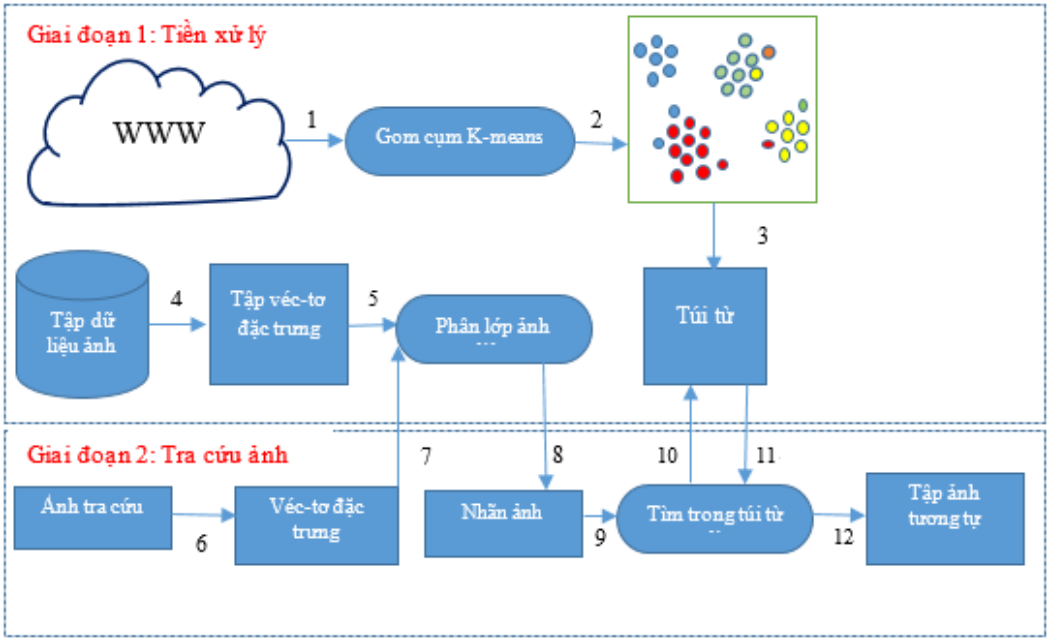
Return ***S***;

End

Mệnh đề 2: Độ phức tạp của thuật toán **CkNN** là $O(n * m)$. Với n số phần tử trong tập véc-tơ đặc trưng F, m là số cụm.

Chứng minh: Gọi n là số véc-tơ đặc trưng trong bộ dữ liệu ảnh. Với mỗi véc-tơ đặc trưng $f_I \in F$ thuộc bộ dữ liệu, thuật toán tiến hành đo khoảng cách Euclide giữa f_I đến m tâm cụm để tìm ra phân lớp cho ảnh đầu vào. Vì vậy độ phức tạp là $O(n * m)$. ■

3.3. Mô hình tra cứu ảnh



Hình 4. Mô hình tra cứu ảnh

Trong Hình 4, (1) thực hiện gom cụm tập ảnh thu thập từ nguồn internet theo phương pháp K-means; (2) kết quả sau khi thực hiện gom cụm là k cụm; (3) xây dựng túi từ dựa vào k cụm; (4) với mỗi ảnh trong CSDL, tiến hành rút trích đặc trưng; (5) phân lớp tập véc-tơ đặc trưng; (6) rút trích đặc trưng của ảnh tra cứu; (7) phân lớp ảnh tra cứu này; (8) nhãn kết quả; (9) tìm tập ảnh tương tự; (10) dựa vào nhãn kết quả, tìm trong túi từ; (11) trả về túi từ tương ứng với nhãn cần tìm; (12) trả về tập ảnh tương tự với ảnh tra cứu đầu vào.

3.4. Thuật toán tra cứu ảnh

Đầu tiên, nhóm tác giả xây dựng các túi từ thị giác cho tập dữ liệu ảnh đầu vào dựa trên véc-tơ đặc trưng và thuật toán K-Means. Với mỗi ảnh truy vấn đầu vào, thuật toán phân lớp k-NN được thực hiện để phân lớp ngữ nghĩa. Dựa vào lớp ngữ nghĩa tìm được, danh sách ảnh tương tự được trích xuất từ cấu trúc túi từ. Thuật toán tra cứu ảnh (CBIR) được mô tả như sau:

Thuật toán CBIR
Đầu vào: Véc-tơ đặc trưng *f* của ảnh tìm kiếm *I*, tập véc-tơ đặc trưng F, Túi từ thị giác.
Đầu ra: Tập ảnh tương tự S_I

Begin

$S_I = \emptyset$;

S = **CkNN** (f_I ,*F*,*k*, θ);

Foreach ($\Omega_i \in \Omega$) **do**

If ($\Omega_i.Label = S$) **Then**

$S_I = S_I \cup \Omega_i$;

EndForeach

Return S_I;

End.

Mệnh đề 3: Độ phức tạp của thuật toán **CBIR** là $O(n * m * k)$. Với *n* số phần tử trong tập véc-tơ đặc trưng F, *m* cụm và *k* túi từ.

Chứng minh: Với véc-tơ đặc trưng f_I của ảnh đầu vào, hệ thống sử dụng thuật toán **CkNN** tiến hành phân lớp ảnh dựa vào tập véc-tơ đặc trưng *F* với độ phức tạp $O(n * m)$ (chứng minh trên). Sau đó hệ thống duyệt qua *k* túi từ để tìm kiếm tập ảnh tương tự với ảnh đầu vào. Vì vậy, độ phức tạp là $O(n * m * k)$. ■

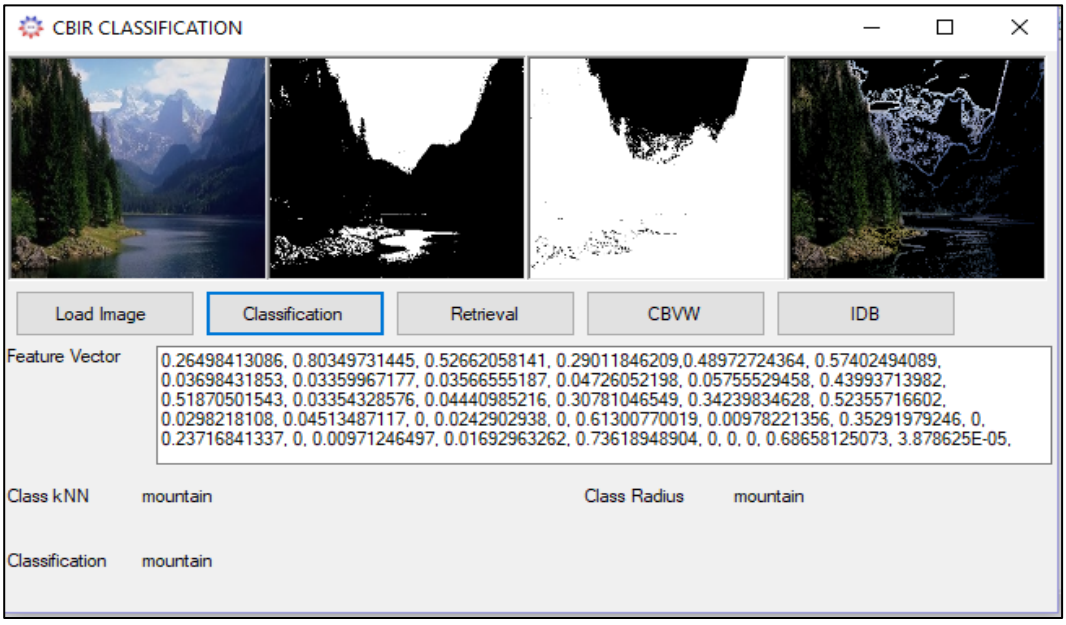
4. THỰC NGHIỆM

4.1. Mô tả thực nghiệm

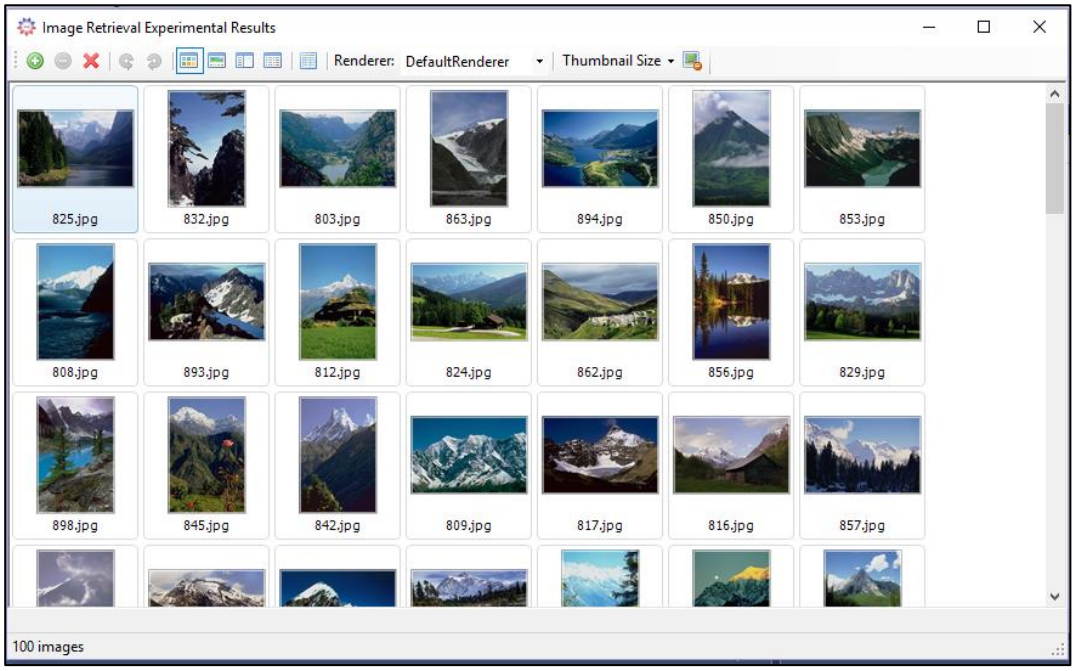
Hệ thống được thử nghiệm trên bộ dữ liệu ảnh COREL (1000 ảnh) (được lấy từ nguồn www.corel.com), trong đó bộ ảnh được chia thành 10 phân lớp, gồm các phân lớp đối tượng và ảnh phong cảnh: Beach, Bus, Castle, Dinosaur, Elephant, Flower, Horse, Meal, Mountain, People. Trong thực nghiệm này, nhóm tác giả sẽ lần lượt truy vấn từng ảnh trên bộ dữ liệu COREL và đánh giá hiệu suất phân lớp cũng như thời gian truy vấn ảnh tương tự.

Thực nghiệm gồm 2 giai đoạn: (1) giai đoạn tiền xử lý rút trích tập các véc-tơ đặc trưng từ bộ dữ liệu ảnh và xây dựng các túi từ bằng thuật toán CBVW; (2) giai đoạn tra cứu và tìm tập các ảnh tương tự thông qua kỹ thuật *k*-NN kết hợp BoVW. Các ứng dụng thực nghiệm được xây dựng trên nền tảng dotNET Framework 3.5, ngôn ngữ lập trình C#. Thực nghiệm trên máy PC CPU Intel (R) Core i5-2430M CPU @2.40GHz, 4.0 GB RAM, hệ điều hành Windows 7 Pro 64 bit.

Trong Hình 5, các véc-tơ đặc trưng được trích xuất từ các vùng của ảnh, với các đặc trưng này bao gồm vị trí, màu sắc, chu vi đối tượng, diện tích đối tượng. Độ tương tự được thực hiện dựa trên khoảng cách trung bình của các véc-tơ đặc trưng theo từng nhóm đặc tính và được tính toán theo độ đo Euclide. Mỗi hình ảnh được phân lớp dựa trên thuật toán k-NN đã được đề xuất để tìm ra các tập ảnh tương tự theo phân lớp đó. Hình 6 mô tả một kết quả truy vấn ảnh dựa trên kết quả phân lớp của Hình 5, với các hình ảnh trong Hình 6 được trích xuất từ một túi từ thị giác bao gồm các hình ảnh tương ứng với các ngữ nghĩa phân lớp ban đầu.



Hình 5. Giao diện chính ứng dụng phân lớp và tra cứu ảnh



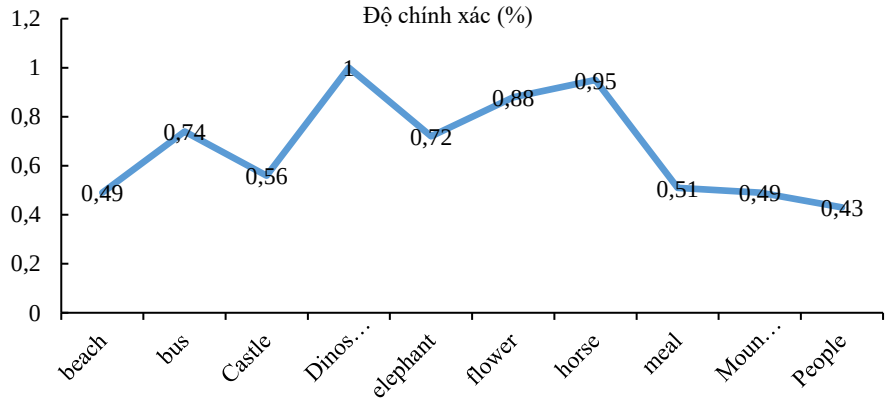
Hình 6. Một kết quả tra cứu ảnh

4.2. Đánh giá kết quả thực nghiệm

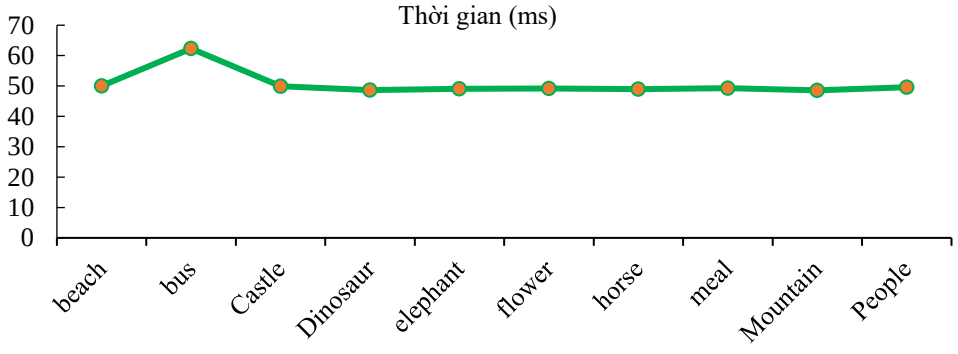
Kết quả thực nghiệm được đo đạc trực tiếp từ chương trình dựa trên bộ ảnh kiểm thử theo từng phân lớp của bộ ảnh COREL. Sau đó, các giá trị thực nghiệm được thống kê và tính giá trị trung bình, trong đó độ chính xác và thời gian truy vấn được mô tả trong Bảng 1 và 2. Kết quả thực nghiệm cho thấy phương pháp phân lớp và truy vấn ảnh đạt được độ chính xác cao và thực hiện với tốc độ tương đối nhanh; với thời gian thực hiện trung bình thử nghiệm trên bộ ảnh COREL là 50,54 ms (milisecond), độ chính xác trung bình là 67,7%.

Bảng 1. Độ chính xác và thời gian thực hiện thuật toán trên bộ ảnh COREL

Phân lớp ảnh	Độ chính xác (%)	Thời gian (ms)
Beach	49	49,98
Bus	74	62,30
Castle	56	49,91
Dinosaur	100	48,65
Elephant	72	49,05
Flower	88	49,12
Horse	95	48,98
Meal	51	49,30
Mountain	49	48,56
People	43	49,63
Trung bình	67,70	50,54



Hình 7. Biểu đồ thể hiện độ chính xác trên bộ ảnh COREL



Hình 8. Biểu đồ thể hiện thời gian thực hiện thuật toán trên bộ ảnh COREL

Hình 7 và Hình 8 mô tả độ chính xác phân lớp trung bình và thời gian truy vấn theo ms (milisecond), trong đó trục ngang của đồ thị mô tả tên phân lớp của bộ ảnh COREL, trục đứng của đồ thị lần lượt mô tả độ chính xác và thời gian truy vấn ảnh.

Qua số liệu về thời gian thực thi và độ chính xác của thuật toán trên bộ dữ liệu COREL (Bảng 1, 2) và Hình 3, 4 cho thấy độ chính xác trên bộ Dinosaur, Horse, Flower, Bus khá cao (100%, 95%, 88%, 74%), tức là phương pháp truy vấn rất khả thi cho các hình ảnh đối tượng. Tuy nhiên, trên các bộ Beach, People, Mountain còn hạn chế, tức là trong các bộ ảnh về phong cảnh thì phương pháp truy vấn đã đề xuất cần phải được cải tiến. Thời gian thực thi trung bình của thuật toán trên các bộ là khá tốt.

Bảng 2. So sánh độ chính xác giữa các phương pháp trên bộ dữ liệu CIFAR-10

Phương pháp	Độ chính xác trung bình (MAP)
Imran M., 2014 [16]	0,5890
Huneiti A., 2015 [18]	0,5588
Shrinivasacharya P., 2015 [19]	0,6537
Phương pháp của chúng tôi	0,6670

Nhóm nghiên cứu Imran và cộng sự (2014) sử dụng bố cục màu MPEG-7 và kết cấu làm cơ sở để trích xuất đặc trưng [16]. Tuy nhiên, việc sử dụng chủ yếu đặc trưng màu sắc để so sánh dẫn đến kết quả truy vấn (P@10) chỉ đạt 58,9%. Tại thời điểm truy vấn, hình ảnh không được phân lớp, nên những bộ ảnh như Bus, Horse chỉ đạt 34% và 53%, trong khi kết quả của chúng tôi là vượt trội hơn nhiều với độ chính xác lần lượt là 74% và 95%.

Nhóm nghiên cứu của Huneiti và cộng sự (2015) thực hiện phân nhóm hình ảnh dựa trên hệ số màu trước khi thực hiện so sánh véc-tơ đặc trưng kết cấu của hình ảnh truy vấn, do đó nhóm ảnh có sự phân biệt về màu sắc rõ ràng như Flower cho độ chính xác khá cao (82,8%), trong khi đề xuất của nhóm tác giả có kết quả cho bộ Flower là 88%. Bộ ảnh về Dinosaur có màu sắc tương đồng thì việc phân biệt về kết cấu của Huneiti và cộng sự không cho kết quả cao (52,6%), và đề xuất trong bài báo này của nhóm tác giả là 100%. Qua đó cho thấy, những đề xuất của nhóm tác giả trong bài báo này về việc truy vấn hình ảnh dựa vào phân lớp hình ảnh theo túi từ có kết quả tốt hơn nhóm nghiên cứu Huneiti và cộng sự.

Khi so sánh với các nghiên cứu kể trên cho thấy kết quả nghiên cứu của nhóm tác giả là hiệu quả.

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, nhóm tác giả đã đề xuất một cải tiến thuật toán *k*-NN và mô hình truy vấn ảnh dựa trên túi từ nhằm phân lớp dữ liệu để tạo ra các phân loại ngữ nghĩa cho hình ảnh, xây dựng cấu trúc túi từ thị giác để tìm kiếm hình ảnh tương tự. Kết quả thực nghiệm trên bộ dữ liệu ảnh COREL được đánh giá và so sánh với các công trình khác trên cùng một tập dữ liệu ảnh đã cho thấy phương pháp đề xuất là hiệu quả. Kết quả thực nghiệm cho thấy, thời gian truy vấn và độ chính xác phân lớp ảnh của bài toán tìm kiếm ảnh là khả thi. Chúng tôi đã cải tiến thuật toán k-NN và kết hợp với mô hình túi từ để giải quyết bài toán là một phương pháp hiệu quả và có thể áp dụng được trong các hệ truy vấn ảnh. Hướng cải tiến tiếp theo là nhóm tác giả sẽ trích xuất đặc trưng phù hợp với hình ảnh phong cảnh, đồng thời truy vấn ngữ nghĩa của các phân lớp hình ảnh trên Ontology để tạo ra các ngữ nghĩa liên quan với các đối tượng trên ảnh.

TÀI LIỆU THAM KHẢO

1. Patrizio A. - IDC: Expect 175 zettabytes of data worldwide, Network World, Dec 3, 2018. https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html.
2. David R., John G., John R. - The digitization of the world: from edge to core, sponsored by Seagate, IDC Technical Report (2018). https://www.seagate.com/as/en/our-story/data-age-2025/.
3. Deloitte, Photo sharing: trillions and rising, Deloitte Touche Tohmatsu Limited, Deloitte Global, 2016.
4. Muneesawang P., Zhang N., Guan L. - Multimedia database retrieval: Technology and applications, Springer, New York Dordrecht London (2014).
5. Xie X., Cai X., Zhou J., Cao N., Wu Y. - A semantic-based method for visualizing large image collections, IEEE Transactions on Visualization and Computer Graphics **25** (7) (2019) 2362-2377.
6. Deligiannidis L., Arabnia H.R. - Emerging trends in image processing, computer vision, and pattern recognition, Elsevier, USA: Morgan Kaufmann, Waltham, MA 02451 (2015).
7. Liu Y., Zhang D., Lu G., Ma W.Y. - A survey of content-based image retrieval with high-level semantics, Pattern Recognition Journal **40** (2007) 262 - 283.
8. Alzu'bi A., Amira A., Ramzan N. - Semantic content-based image retrieval: A comprehensive study, J Vis Commun Image Represent **32** (2015) 20-54.
9. Erwin Fachrurrozi M., Ahmad F., Bahardiansyah R.S., Rachmad A., Anggina P. - Content based image retrieval for multi-objects fruits recognition using k-means and k-nearest neighbor, 2017 International Conference on Data and Software Engineering (ICoDSE), Palembang (2017) 1-6.
10. Kumar M., Payal C., Naresh K. G. - An efficient content based image retrieval system using BayesNet and K-NN, Multimedia Tools and Applications **77** (16) (2018) 21557-21570.
11. Shichao K., Lihui C., Xinwei Z., Yigang C., Zhenmin Z. Hengyou W. - A supervised learning to index model for approximate nearest neighbor image retrieval, Signal Processing: Image Communication **78** (2019) 494-502.
12. Yanchun M., Wing X., Yongjian L., Shengwu X. - A weighted KNN-based automatic image annotation method, Neural Computing and Applications (2019) 1-12.
13. Alqasemi F. A., Alabbasi H.Q., Sabeha F., Alawadhi A., Kahlid S., Zahary A. - Feature selection approach using KNN supervised learning for content-based image retrieval, 2019 First International Conference of Intelligent Computing and Engineering (ICOICE), Hadhramout, Yemen (2019)1-5.
14. Zhang H., Serkan K., and Moncef G. - A k-nearest neighbor multilabel ranking algorithm with application to content-based image retrieval, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA (2017) 2587-2591.
15. Xiaohui S., Zhe L., Jonathan B., Ying W. - Spatially-constrained similarity measurefor large-scale object retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence **36** (6) (2013) 1229-1241.

16. Imran M., Hashim R., Abd Khalid N. E. - Content based image retrieval using MPEG-7 and histogram, In: Herawan T., Ghazali R., Deris M. (Eds.) - Recent Advances on Soft Computing and Data Mining, Advances in Intelligent Systems and Computing **287**, Springer International Publishing, Switzerland (2014) 453-465.
17. Li D., Mooi C.C. - A novel unsupervised 2-stage k-NN re-ranking algorithm for image retrieval, IEEE International Symposium on Multimedia (ISM), Miami, FL (2015) 160-165.
18. Huneiti A., Daoud M. - Content-based image retrieval using SOM and DWT, Journal of software Engineering and Applications **8** (2) (2015) 51-61.
19. Shrinivasacharya P., Sudhamani M. V. - Content based image retrieval using self organizing map, In: Proceedings of the Fourth International Conference on Signal and Image Processing (2015) 535-546.
20. Zhang H., Serkan K., and Moncef G. - A k-nearest neighbor multilabel ranking algorithm with application to content-based image retrieval, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA (2017) 2587-2591.
21. Shuang J., Lin M., Xuezhi T., Danyang Q. - Bag-of-visual words based improved image retrieval algorithm for vision indoor positioning, IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium (2020) 1-4.

ABSTRACT

A METHOD OF CLASSIFICATION FOR K-NN BASED IMAGE RETRIEVAL

Huynh Thi Chau Lan*, Le Huu Ha, Nguyen Hai Yen
Ho Chi Minh City University of Food Industry
*Email: *lanhtc@hufi.edu.vn*

In this paper, a stratified data approach was applied to a similar image search problem through a bag vision feature from BoVW (Bag of Visual Words). The classification method is based on the k-NN (k-Nearest Neighbor) algorithm with the input data being a feature vector of the image. From an initial image data set, we construct a bag of visual words to stores images that are substantially similar in content. After classifying the input image by the k-NN method, a set of similar images is extracted from BoVW. In the k-NN method, in addition to k nearest neighbors, a radius θ is used to statistically classify the image. Each BoVW links to other word bags through its representative semantic class. Experiments were built on COREL image database (1,000 images) to evaluate the accuracy and compare with other related works on the same data set. According to empirical results, our recommendations are effective and can be applied in various multimedia systems.

Keywords: k-NN (k-Nearest Neighbor), classification, bag of words, similar image, similarity measure.