Journal of Information Hiding and Multimedia Signal Processing ©2020 ISSN 2073-4212
Ubiquitous International Volume **11**, Number **3**, September **2020**

Improve The Efficiency Of Content-based Image Retrieval Through Incremental Clustering

Quynh Dao Thi Thuy
Faculty of Information Technology
Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
quynhdao.ptit@gmail.com

Quynh Nguyen Huu
Faculty of Computer Science and Engineering
Thuyloi University, 175 Tayson, HaNoi, Viet Nam
quynhnh@tlu.edu.vn

Phuong Nguyen Thi Lan
Thai Nguyen University - Lao Cai Campus
phuongntl@tnu.edu.vn

Tao Ngo Quoc
Institute of Information Technology
Vietnam Academy of Science and Technology, Hanoi, Viet Nam
nqtao@ioit.ac.vn

Minh-Huong Ngo
Institute of sciences of Digital, Management and Cognition
University of Lorraine, France
ngominhhuong95@gmail.com

Received September 2019; revised February 2020

Abstract. *In recent years, the relevant feedback approach, widely used in image retrieval methods, has improved the accuracy of image retrieval. However, the retrieval time of the methods is still high because they have to re-cluster the entire feedback image set. In this paper, to reduce the time taken for image retrieval, we propose a method of retrieving images with incremental clustering, named IRIC (Image retrieval method using Incremental clustering), the method does not re-cluster all of the feedback images of the user. The experiments were performed on a set of 10,800 images and the results demonstrate that the proposed method improves the performance of the system.*
Keywords: Content-based image retrieval, Relevant feedback, Incremental clustering

1. **Introduction.** Recently, image retrieval has attracted the attention of many researchers in the computer science community. With the availability of digital image acquisition devices, the size of the digital image set increased rapidly. The content-based image retrieval system becomes the key to the effective use of this digital resource. In a typical CBIR system, low-level visual features (colors, textures, and shapes) are automatically extracted for indexing and image description purposes. To search for desired images, the user takes a sample image and the system returns a set of similar images based on extracted features. When the image retrieval system presents a set of images that are considered similar to the query, users can feedback their opinions to the image retrieval system through marking images that are relevant to the given query image. The system adjusts queries based on images that are marked by the user. Relevant feedback in the CBIR does not require users to provide accurate initialization query images that improve the accuracy of the CBIR system through user-feedback images.

During the implementation of related feedback, new query points in the feature space are calculated and the distance function is adjusted. Works in a single-point approach represent a new query with a single point and change the weights of specific components to find an optimal query point and an optimal distance function [1]. In this approach, a single point is calculated by the weighted average of all related images in the feature space. The multi-point approach represents a new query with multiple points to determine the shape of the border [2]. This approach calculates new query points through a method of clustering related images that users provide. Assuming that the relevant images are mapped to similar points of similar measurement, a wide border is built to cover all query points and the system looks for images similar to these queries. However, if the feature space and distance function are very different from the user's perception, the relevant images are mapped to discrete regions of any shape in the feature space. That is, related images can be ranked under images that are searched according to a given query. To converge quickly to user information needs, the system will find similar images with any query point. A query that looks for images similar to any query point is called a separate query. A complex image query is represented by separate regions because semantically related images can be scattered in some visual areas rather than one region. Due to the superiority of a separate query, in this paper, we follow a separate query approach (multi-point query).

The above image retrieval systems have limitations that are a slow retrieval time. The reason for this is illustrated by the example in Figure 1. In this figure, a circular dot represents a feedback image, the ellipses represent a cluster.

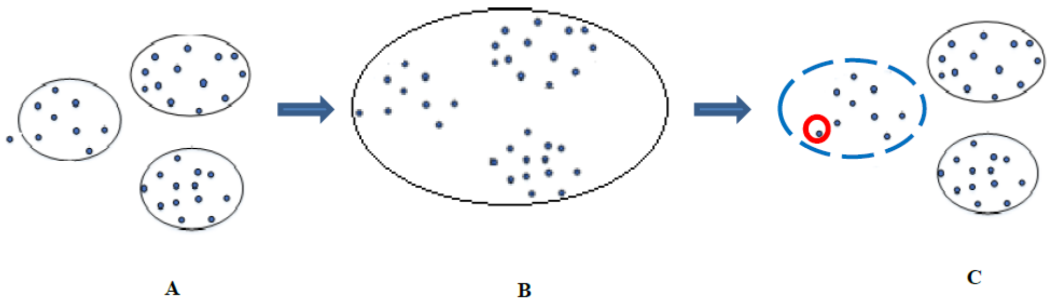

FIGURE 1. Illustrating the re-clustering process

Figure 1 illustrates the process of defining new query points with the re-clustering of existing methods. First, at the previous feedback loop, the user feedback images are clustered into three clusters 1, 2 and 3. At the next iteration, the user marks a related image (see Block A). Next, all images are combined to perform the re-clustering process (see Block B). Implementing some clustering algorithm for images of block B, we will get three clusters 1, 2 and 3 of block C. The example in Figure 1 shows that, for each feedback loop, we have to re-cluster all the feedback images. This makes the search speed of the methods already very slow.

In this paper, we propose a content-based retrieval method through incremental clustering. Instead of re-clustering all the feedbacks of two iterations, our method will determine the probability that an image belongs to which cluster is the largest to add to that cluster. Because the dimension of the feature space in the image retrieval is often very high (the number of dimensions can be from tens to hundreds in most cases), so estimating the model directly in the high dimensional feature space is easy to fail. Therefore, our proposed method also uses dimensional reduction to map the original feature space to a low-dimensional feature space. In addition, our proposed method applies the optimal distance function to improve accuracy.

The rest of the article is organized as follows. In Part 2, we briefly present research on image retrieval methods using related feedback. Part 3 describes in detail the proposed method. Part 4 describes our performance experiments and discusses the results. Finally, we conclude Part 5.

2. **Related Works.** Methods based on the traditional approach [28, 26, 24, 27, 35, 36, 29], image retrieval based on the similarity of low-level visual features often, have poor performance. To solve this problem, several techniques using related feedbacks [1, 3, 4, 5, 7, 8, 22, 23, 24, 25, 30, 31, 32, 33, 34, 35] have been proposed. These works link semantic concepts and low-level imaging features through user feedback.

The query point movement approach has been applied to image retrieval systems such as MARS [3], MindReader [5] and the effective method of shifting query points [7]. These systems represent the query as a single point in the feature space and try to move that point in the direction of positive points and away from the negative points. This idea originated from Rocchio's algorithm [6], which has been successfully used in document lookups. The methods using this approach have the limitation that it is impossible to retrieve semantically related images scattered throughout the entire visual feature space.

To overcome the limitations of single-point queries, methods that use the multi-point query movement approach (query expansion approach) have been introduced. MARS's query expansion approach [38] builds local clusters for related points. In MARS, all local clusters are merged to create a wide border covering all query points. On the other hand, the point query movement approach [3, 5] also ignores the clusters and considers all related points to be equivalent. These two approaches can generate hyper ellipsoids or convex shapes using clusters in feature spaces to cover all query points for simple queries. However, both approaches fail to identify the appropriate regions for complex queries. [8] presented FALCOM, the entire distance model to facilitate the learning of concave and disjoin query points in vector space as well as in any measurement space. However, the proposed distance function depends on specific empirical knowledge and this model assumes that all the relevant points are query points. Y. Chen et al. [9] also adopted this approach to use it in multiple seed queries but they used them to extend the boundaries around the best query and still search primarily in a single region. of feature space. Quynh et al. [37] proposed the SRIR method, which has the advantage of not requiring users to provide many appropriate query images to represent their information needs, determining the semantic importance of each query (associated with distance) and importance according to each feature to improve accuracy. In addition, the SRIR method reduces the burden on users in the process of returning semantic images scattered throughout the featured space with high accuracy. The multi-point approach has to cluster the related images, so it takes a lot of time to retrieval.

3. **Proposed Method.** As mentioned in the previous section, most image search methods using a multi-point approach, which uses clustering techniques, must recluster the feedback image set. This leads to the time of the retrieval system is high. In this section, we propose an image retrieval method with incremental clustering (IRIC) without having to re-cluster related image set, to reduce image retrieval time.

3.1. **Image retrieval scheme of the proposed method.** The IRIC method is described by the schema in Figure 1. With the user-supplied query image, the method of performing the initialization query to obtain the initialization result set. After the user feedback on the initialization result set, we get the feedback set. A clustering process with dimensional reduction is performed to obtain image clusters. These image clusters are used to create an initial training set and a list of query points. Based on the existing image clusters, the method of calculating the centers of clusters and taking these centers as the corresponding query points. Next, the method implements a multipoint query to obtain the result set. If the retrieval process stops, this result set is the final result set. In contrast, the retrieval process is continued as follows:

With the first feedback loop: Users make feedback to get feedback set. The feedback set and the initial training set are used for incremental clustering. After performing the incremental clustering process, the method will return a list of clusters. Next, we compute cluster centers to obtain a list of cluster centers. These cluster centers are taken as the corresponding query points. Performing a multipoint query, we will obtain the result set. In a situation where the user stops responding, the method will return the final result set. In contrast, the second feedback iteration is carried out, which is similar to the first, with the only difference being that the incremental clustering process does not reuse the initial training set. From the third feedback loop onwards, do the same as the second iteration. Subsequent feedback iterations are performed in the same way as the second iteration.

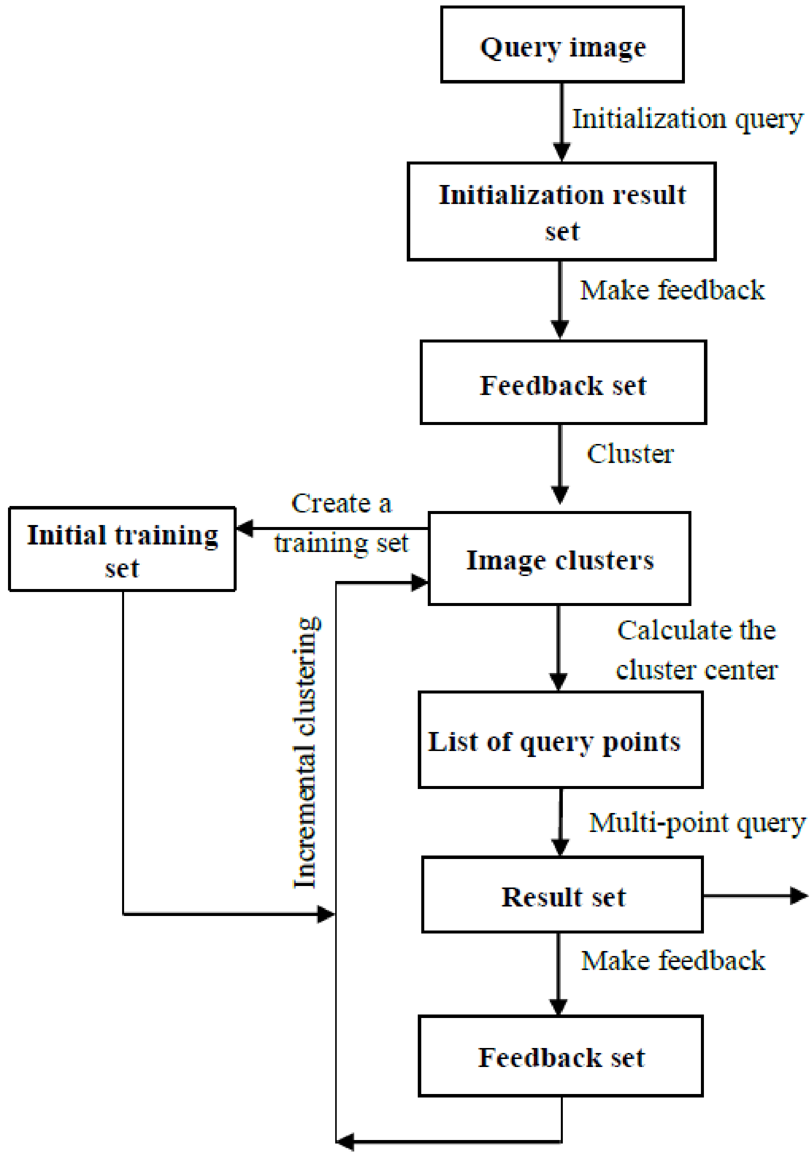

FIGURE 2. Image retrieval scheme with incremental clustering

3.2. **Clustering with dimensional reduction.** Data representation is an important first step to solving a clustering problem. In computer vision, there are two types of representations that are widely used: geometric and graphical representation [11]. In this paper, we choose the second type of representation because it applies to nonmetric distances.

In a weighted undirected graph G, the nodes $V = \{s_1, s_2..., s_n\}$ represent images, the edges $E = \{(s_i, s_j) : s_i, s_j \in V\}$ represent the relationship of each pair of nodes, and a non-negative weight a_{ij} of an edge (s_i, s_j) indicates the similarity between the s_i and s_j nodes. The elements of the affinity matrix A [12] are calculated by the formula (1):

$$(a_{ij} = e^{\frac{-\|s_i - s_j\|^2}{2\sigma^2}}) \quad (i \neq j), a_{ii} = 0) \tag{1}$$

Here the parameter σ^2 controls how the affinity a_{ij} decreases with the distance between s_i and s_j. The value a_{ij} between two images is "high" if the two images are very similar.

Based on the representation of the image set by a weighted undirected graph, we perform this graph partition. We present a clustering problem as a graphing partitioning problem. In a graph partition, nodes are organized into groups so that the similarity within the group is high, and/or similarity between groups is low.

Among many methods of partitioning, Spectral Graph Partitioning [13, 14] has been successfully applied to many areas of computer vision [15, 13, 14]. There are two methods in the Spectral Graph Partitioning: the method uses k separate vectors [16] and the method uses a separate vector [13]. The first method calculates the k-way partition directly, the second method uses a single vector at a time and implements a recursive mechanism. In this paper, we use the first method because the method of obtaining a k-way partitioning directly can produce much better partitionings than the one that calculates a k-way partitioning via the recursive mechanism.

Below, we briefly present the method of A. Y. Ng et al. (See more details in [16]).

First, from n points (images), we construct affinity matrix A by formula (1). Construct diagonal matrix D in which the element (i, i) is the sum of the ith row of matrix A. D is a diagonal matrix with:

$$(D_{ii} = \sum_{j=1,\dots,n} a_{ij}) \tag{2}$$

The standardized Laplace matrix is calculated by (3)

$$(L = D^{\frac{-1}{2}} A D^{-1}) \tag{3}$$

Find k largest eigenvectors $x_1, x_2, \dots x_k$ of the matrix L, where $x_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$, $x_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n}), \dots, x_k = (x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn})$ and $X = [x_1{}^T, x_2{}^T, \dots, x_k{}^T] \in R^{n\ x\ k}$.

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

Construct matrix Y from X by normalizing each line of X according to the unit length of matrix Y:

$$Y_{ij} = \frac{X_{ij}}{(\sum_j X_{ij}^2)^{\frac{1}{2}}} \tag{4}$$

Each row of the matrix Y is considered as a point in the k-dimensional vector space. Thus, we have n points in space R^k, grouping $(y_i)_{i=1\dots n}$ in space R^k into k clusters $C_1, C_2, \dots, C_k$ through K-Means. Then we assign a point s_i to cluster j if and only if the ith row of the matrix Y corresponds to cluster j.

$$
\begin{array}{c|ccccc}
 & & & & & \\
\hline
\boldsymbol{y_1} & y_{11} & y_{12} & y_{13} & \cdots & y_{1k} \\
\boldsymbol{y_2} & y_{21} & y_{22} & y_{32} & \cdots & y_{2k} \\
\boldsymbol{y_3} & y_{31} & y_{32} & y_{33} & \cdots & y_{3k} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{y_k} & y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nk}
\end{array}
$$

Algorithm 1 below, named CISE (Clustering Images Set using Eigenvectors), is a clustering algorithm with dimensional reduction. This algorithm groups images into k groups.

Algorithm 1 CISE

Input: - Image set S=$\{s_1, s_2, \ldots, s_n\}$ with $s_i \in R^n$
-Number of clusters: k
Output: k clusters: $C_1, C_2, \ldots, C_k$.

1. Construct an affinity matrix
for $i \leftarrow 1$ to n **do**
  for $j \leftarrow 1$ to n **do**
    if $i \neq j$ **then**
      $a_{ij} \leftarrow \exp\left(\frac{-\|s_i - s_j\|^2}{2\sigma^2}\right)$
    else
      $a_{ij} \leftarrow 0$
2. Construct a diagonal matrix and a Laplace matrix L
for $i \leftarrow 1$ to n **do**
  $d_{ii} \leftarrow \sum_{j=1,\ldots,n} a_{ij}$
$L \leftarrow D^{-1/2} A D^{-1/2}$
3. Find k largest eigenvectors $x_1, x_2, \ldots x_k$ **of the Laplace matrix L**
for $i \leftarrow 1$ to k **do**
  $x_i \leftarrow Largest\_eigen\_vectors(L)$
$X \leftarrow [{x_1}^T, {x_2}^T, \ldots, {x_k}^T]$
4. Construct a matrix Y from X
for $i \leftarrow 1$ to n **do**
  for $j \leftarrow 1$ to k **do**
    $y_{ij} \leftarrow x_{ij} / (\sum_k x_{ik}^2)^{1/2}$
$Y \leftarrow [y_1, y_2, \ldots, y_k]$
5. Cluster through K-Means
$P \leftarrow \emptyset$
for $i \leftarrow 1$ to n **do**
  $p_i \leftarrow y_i$
  $P \leftarrow P \cup p_i$
K-Mean(P)
6. Assign s_i **to the clusters**
for $i \leftarrow 1$ to n **do**
  if $p_i \in (C_j)_{i=1,..k}$ **then**
    $C_j \leftarrow C_j \cup s_i$
return $C_1, C_2, \ldots, C_k$

3.3. **The proposed incremental clustering algorithm.** There are many clustering algorithms such as K-means, K-medoid, etc. which are used in image retrieval methods [17, 18]. However, when a new image is added, the methods must recluster all the images. Therefore, these methods do not suit the case of online requests, for example, in the case of a small feedback set but require immediate clustering and many images still need to be added and clustered next. The algorithms that satisfy this online case are called "incremental" or "incremental clustering". In incremental clustering algorithms, determining the cluster for an object is the most important task. Below, we describe our proposed incremental clustering algorithm.

Assume that the data has a Gauss distribution. In this algorithm, we treat each cluster as a group. When training, we will estimate the center of each group and the covariance matrix. The task of determining the cluster of an object is considered as the problem of finding an estimate $P\left(Y|X\right)$ such that: for an input x_0, its cluster label will be identified by:

$$\hat{y}_0 = argmax_y P(Y = y|X = x_0) \tag{5}$$

However, $P\left(Y|X\right)$ is difficult to calculate, so instead of calculating $P\left(Y|X\right)$, we will estimate through $P\left(X|Y\right)$ and $P\left(Y\right)$. According to Bayes' rule, where i is the label of the group, we have the formula:

$$P\left(Y=i|X=x\right) = \frac{P\left(X=x|Y=i\right)P\left(Y=i\right)}{P\left(X=x\right)} = \frac{P\left(X=x|Y=i\right)P\left(Y=i\right)}{\sum_j P\left(X=x|Y=j\right)P\left(Y=j\right)} \tag{6}$$

Assume that $P\left(X=x|Y=i\right)$ is a multivariate normal distribution with a density:

$$f_i\left(x\right) = \frac{1}{\left(2\pi\right)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}} e^{\frac{-1}{2}\left(x-\mu_i\right)^T|\sum|^{-1}\left(x-\mu_i\right)} \tag{7}$$

Where:
μ_i : Mean of the inputs for group i
$\sum$: Covariance matrix (common to all groups)
Suppose that we know:

$$P\left(Y=i\right) = \pi_i \tag{8}$$

$$= \frac{\#\{j; y_j = i\}}{N} \tag{9}$$

Note: formula (9) is the ratio of the training samples of group i to the total number of training samples.

At this point, we obtain the formula:

$$P\left(Y=i|X=x\right) = \frac{f_i\left(x\right)\pi_i}{P\left(X=x\right)} \tag{10}$$

Since the denominator in (10) is independent of i, we can consider it a constant C and obtain the formula:

$$P\left(Y=i|X=x\right) = C \times f_i\left(x\right)\pi_i \tag{11}$$

Replacing $f_i\left(x\right)$ from (7) into (11), we get:

$$P\left(Y=i|X=x\right) = \frac{C\pi_i}{\left(2\pi\right)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}} e^{\frac{-1}{2}\left(x-\mu_i\right)^T|\sum|^{-1}\left(x-\mu_i\right)} \tag{12}$$

Because $\left(2\pi\right)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}$ in (12) does not depend on i, we set $\frac{C}{\left(2\pi\right)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}}$ equal to the constant C' and we have:

$$P\left(Y=i|X=x\right) = C'\pi_i e^{\frac{-1}{2}\left(x-\mu_i\right)^T|\sum|^{-1}\left(x-\mu_i\right)} \tag{13}$$

and take the natural of both sides of (13), we get:

$$\ln P(Y = i\,|X = x\,) = \ln C' + \ln \pi_i - \frac{1}{2}(x - \mu_i)^T \sum\nolimits^{-1} (x - \mu_i) \tag{14}$$

In (14), the $\log C'$ value of the right side is true for all groups i, so we are only interested in:

$$\ln \pi_i - \frac{1}{2}(x - \mu_i)^T \sum\nolimits^{-1} (x - \mu_i) = \ln \pi_i - \frac{1}{2}\left[x^T \sum\nolimits^{-1} x + \mu_i^T \sum\nolimits^{-1} \mu_i\right] + x^T \sum\nolimits^{-1} \mu_i \tag{15}$$

Thus, our goal is to maximize the formula (15) in i.

In (15), since $x^T \sum^{-1} x$ is independent of i, we consider it a constant C'' and (15) transformed into

$$C'' + \ln \pi_i - \frac{1}{2}\mu_i^T \sum\nolimits^{-1} \mu_i + x^T \sum\nolimits^{-1} \mu_i \tag{16}$$

Ignoring the constant C'', we have the objective function:

$$\delta_i(x) = \ln \pi_i - \frac{1}{2}\mu_i^T \sum\nolimits^{-1} \mu_i + x^T \sum\nolimits^{-1} \mu_i \tag{17}$$

With an input x, we predict its label as i when $\delta_i\,(x)$ is the largest.

Algorithm 2 below, named INC - Incremental Clustering, determines which cluster a new object x_0 belongs to. The INC algorithm has the input of a training set D and an image x_0, In the processing section, the algorithm calculates the value $\delta_i\,(x_0)$ and takes the value i that has $\delta_i\,(x_0)$ as the maximum. The output of the algorithm is cluster i that contains the image x_0,

---

Algorithm 2 INC

Input: - D={ (x_i, y_i) /i=1,. . .,N; $y_i \in \{1, \ldots g\}$ } : training set

- x_0 : image

Output: - i : The cluster contains the image x_0

1: Split D into g clusters, which is based on the number of clusters in Y
2: Calculate the average μ_i of each cluster $i \in \{1, \ldots g\}$ and μ of the whole set D
3: Calculate the covariance matrix of group $i \in \{1, \ldots g\}$ and the common covariance matrix
4: Calculate vectors that replace prior probabilities according to (9)
5: Calculate $\delta_i\,(x_0)$, $i = 1, \ldots g$ by the formula (17)
6: **return** $argmax_i \delta_i\,(x_0)$

3.4. **Proposed retrieval algorithm.** Below is a description of image retrieval algorithm IRIC (image retrieval method using Incremental clustering). This algorithm uses the INC increment clustering algorithm.

Image retrieval algorithm using IRIC incremental clustering is performed as follows:

First, the user enters query Q, the algorithm uses the Euclidean distance function d and returns k images through <m, $\{q_1, q_2, ...q_m\}$, *d*, *S*, *k*> to obtain the results of the initialization query Result($Q_{initial}$). Next, on the initialization result set Result($Q_{initial}$), the user selects N images through the function ***Feedback*** $(Result\ (Q_{initial}), N)$ to obtain the set of N related images of the initialization query Relevant($Qj_{initial}$,N). Cluster the related images of the initialization query Relevant($Q_{initial}$,N) into g clusters and save them to X through the function CISE(Relevant($Q_{initial}$,N), g, X) for a training example set D D← $\{(x_i, y_i)/i = 1, ..., N; y_i \in \{1, ..., g\}\}$, Then, for each cluster i, based on each image $x_j^{(i)}$ (j=1,.., n_i), calculate the optimal query point $q^{(i)}$ through procedure **FQM(**$X^{(i)}$, $q^{(i)}$ **).** Based on g optimal query points $q^{(i)}$ and distance function d, the algorithm returns the

Algorithm 3 IRIC

Input: Set of images: $\boldsymbol{S}$
The query images:$Q_{initial}$=$\{q_i/i = 1...m\}$
The number of images returned at each iteration: k
Output: Result set of the optimal query: ***Result(***Q_{opt}***)***

Result($Q_{initial}$) $\leftarrow< m, q,\ d,\ S,\ k >$;
Relevant($Q_{initial}, N$) $\leftarrow$ ***Feedback*** $(Result\,(Q_{initial})\,, N)$;
CISE(Relevant($Q_{initial}$,N), g, X)
$D \leftarrow \{(x_i, y_i)$/i=1, . . .,N; $y_i \in \{1, \ldots g\}\}$
repeat
    for i=1 to g **do**
        $FQM\left(X^{(i)}, q^{(i)}\right)$
    Result(Q_{opt}) $\leftarrow \langle$ g, $\{q^{(1)}$, $q^{(2)}$,...$q^{(g)}\}$, *D, S, k*$\rangle$;
    Relevant(Q_{opt},N) $\leftarrow$***Feedback*** $\left(Result\,(Q_{opt})\,, N',\right)$;
    for j=1 to N **do**
        INC$(D, x_tj \in Relevant(Q_{opt}, N), i)$;
        Add(x_j, $X^{(i)}$
until (User stops responding)
return Result(Q_{opt});

resulting k images on the set S through <g, $\left\{q^{(1)}, q^{(2)}, ..., q^{(g)}\right\}$, *d, S, k*> and assigns to Result(Q_{opt}). On the result set Result(Q_{opt}), users select N' related images via the function ***Feedback*** $(Result(Q_{opt}), N')$ to get the set Relevant(Q_{opt},N'). The algorithm does not cluster all objects so it is necessary to predict which x_j Relevant(Q_{opt},N') (j=1..N') each cluster belongs to $X^{(i)}$ through procedure **INC**(D, x_j Relevant(Q_{opt},N'), i) and perform the addition of x_j to cluster $X^{(i)}$ through procedure **Add**(x_j, $X^{(i)}$). This process is repeated until the user stops responding. The algorithm ends with a set of result images Result(Q_{opt}).

4. Experiments.

4.1. Experimental environment.

Image set for experiment. The experimental image set, which we reorganized from a subset of Corel Photo Gallery, includes: 10,800 images. There are 80 groups in this image set, some of which are: autumn, aviation, bonsai, castle, cloud, dog, elephant, iceberg, primate, ship, stalactite, fire, tiger, train, waterfall,etc. The number of images in each group is about 100. The size of the images is max (width, height) = 120 and min (width, height) = 80.

There are two types of features: color feature and texture feature. The first feature type includes: color histogram (32 in length), color auto correlogram (with a length of 64) and a color moments (with a length of 6). The second feature type includes wavelet transform (length is 40) and gabor wavelet (length is 48).

Ground truth. Ground truth is widely used in the evaluation of CBIR systems, so we also use the Corel subgroup as the ground truth. We consider all images in the same Corel group to be relevant. This consists of 3 columns (titled: Query image ID, Related image ID) and consists of 1,981,320 rows.

4.2. **Execute and evaluate queries.** We select the parameters for the experiment as follows:

All images in the image set are used as query images. We took the accuracy of the proposed method as the average accuracy of all 10,800 query images. The result returned for each query is 100 images. The reason we chose 100 images is that users typically only consider two screen pages to select feedback images and each screen contains 50 images.

For evaluation purposes, we use average precision to compare the effectiveness of other methods. The average precision is the ratio of the number of related images in the returned list to the total number of images returned and it is calculated by the average of all queries. Average accuracy is the main evaluation criteria.

Effective on time. To prove the time-efficiency of the proposed method, we conducted experiments with image retrieval method IRIC_WINC (without using INC incremental clustering algorithm) and image retrieval method IRIC (using incremental clustering algorithm INC).

We also took the Corel set of 10,800 images above as experiments for the two methods. Besides, we also selected 10,800 images in the image set as 10,800 query images. In addition, we took the average execution time of 10,800 query images with three feedback loops as the execution time of each method. As Figure 3 shows, in all three configurations (2, 4 and 8 query points), the execution time of the IRIC method is much lower than the IRIC_WINC method. Especially in the configuration of 8 query points, the execution time of IRIC is lower than IRIC_WINC to 1200 ms. The reason for this is that the IRIC method does not have to re-cluster the feadback image set while IRIC_WINC has to recluster the entire response image set.

FIGURE 3. Retrieval time for two methods (IRIC and IRIC_WINC)

Accuracy of the proposed method. In our experiments, we used the number of query points of 2, 4 and 8 points, respectively. The reason we only use up to 8 query points is because: First, the number of samples for three feedback loops is often not large enough to produce more than eight clusters. Secondly, we want to prove that the accuracy is still high, although the number of query points is not much. We compare the accuracy of our proposed method with two other methods, CRF (Complementary Relevance Feedback) [19] and DSSA (Discriminative Semantic Subspace Analysis) [20].

TABLE 1. The average accuracy of three methods according to the number of query points in the three feedback loops.

Method	**Average accuracy**		
	Two query points	**Four query points**	**Eight query points**
CFR	0.2387	0.3065	0.3199
DSSA	0.3135	0.42658	0.4846
IRIC	**0.3224**	**0.43568**	**0.4895**

In Table 1, the average accuracy of the three methods is CRF, DSSA and the proposed method (IRIC) at levels 2, 4 and 8 query points. In our proposed method, the number of query points is determined by the number of clusters. With 2 query points, the accuracy of the proposed method is higher than that of CRF and DSSA methods, which are 8.37% and 0.89% respectively. In the case of 4 query points, the accuracy of the proposed method is higher than that of the CRF and DSSA methods, which are 12,918% and 0.91% respectively. In the case of 8 query points, the proposed method has higher accuracy than that of the CRF and DSSA methods, which are 16.96% and 0.49% respectively. The accuracy of our proposed method is higher than that of the CRF and DSSA methods because our method has obtained a good data model by reducing the feature dimension and our proposed method has applied the optimal distance function that is described in [39].

To confirm the accuracy of the incremental clustering algorithm, we also evaluate the accuracy of the incremental clustering algorithm through experiments on the Iris data set.

The IRIS dataset includes information about three types of Iris flowers. Three types of flowers include: Iris setosa, Iris virginica and Iris versicolor. Each of these three types has 50 flowers. The data includes 4 attributes: the length, width of the sepals and the length and width of the petals. Each data point in this set is a 4-dimensional vector.

We conducted experiments for three clustering methods on IRIS data set: K-means [21], Spectral [16] and proposed INC algorithm. The efficiency of the K-mean algorithm depends in part on the initialization of cluster centers, so we take its accuracy as the average of three K-mean algorithm runs. The Spectral method was performed once on 150 samples. The INC algorithm is performed in three rounds: the first round is Spectral clustering with 50 samples, the second round is an incremental cluster over 50 samples and the third round is clustering over the remaining 50 samples. The results of the three clustering algorithms are shown in Table 2. As shown in this table, the K-means method has 130 true samples and 20 false samples, the Spectral algorithm has 131 true samples and 19 false samples, the INC algorithm has 132 true samples and 18 false samples. Thus, the correct sample numbers of the three algorithms are equivalent. The accuracy of the INC algorithm is higher than that of the K-means algorithm by 2 samples and is slightly higher than Spectral by 1 sample.

5. **Conclusions.** In this paper, we have proposed an image retrieval method using IRIC incremental clustering, which is designed to reduce time and improve image retrieval accuracy. The IRIC method has the advantage of not having to recluster the entire feedback image set and perform the processing on the dimensional space of the data. Experiments using a large subset of Corel Photo Gallery with 10,800 images demonstrate that the IRIC method outperforms the CRF and DSSA methods in terms of retrieval time and accuracy.

Table 2. Results of three clustering algorithms.

Numerical order	**Clustering algorithms**	**The number of samples correctly clustered**	**The number of samples is incorrectly clustered**
1	K-means	130	20
2	Spectral	131	19
3	INC	132	18

Acknowledgment. The author gratefully acknowledges the many helpful suggestions of the anonymous reviewers during the preparation of the paper. This research was supported by the research support program Offer for senior researchers in 2020 under grant no. NVCC02.01/20-20 and Improve the efficiency of content-based image retrieval through metric learning under grant no. VAST01.07/19-20.

REFERENCES

[1] B. Andre, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, Learning semantic and visual similarity for endomicroscopy video retrieval, *IEEE Transactions on Medical Imaging*, vol. 31, no. 6, pp. 1276-1288, 2012.

[2] R. Arandjelovic and A. Zisserman, Three things everyone should know to improve object retrieval, In *Proc. CVPR*, 2012.

[3] Y. Rui, T. Huang, and S. Mehrotra, Content-based image retrieval with relevance feedback in MARS. *Proceedings of IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct 1997.

[4] K. Porkaew, K. Chakrabarti, Query refinement for multimedia similarity retrieval in MARS, *Proceedings of the 7th ACM Multimedia Conference*, Orlando, Florida, pp. 235-238, 1999.

[5] Y. Ishikawa, R. Subramanya, and C. Faloutsos, Mind Reader: Querying databases through multiple examples. *Proceedings of the 24th VLDBConference*, New York, USA, pp. 218-227, 1998.

[6] J. J. Rocchio, Relevance feedback in information retrieval. *Salton, G. (Ed.), The SMART Retrieval System Experiments in Automatic Document Processing*, Prentice Hall, Englewood Cliffs, NJ, pp. 313-323, 1971.

[7] D. Liu, K. A. Hua, K. Vu, and N. Yu, Fast Query Point Movement Techniques for Large CBIR Systems, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, No. 5, pp. 729-743, 2009.

[8] L. Wu, C. Faloutsos, K. Sycara, and T. R. Payne, FALCON: Feedback adaptive loop for content-based retrieval. *Proceedings of the 26th VLDB Conference*, Cairo, Egypt, pp.297-306, 2000.

[9] Y. Chen, J. Z. Wang, and R. Krovetz, An unsupervised learning approach to content-based image retrieval, *Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Paris, 2003.

[10] X. Jin, J. C. French, Improving Image Retrieval Effectiveness via Multiple Queries, *Multimedia Tools and Applications*, vol. 26, pp. 221-245, 2005.

[11] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, Classification with nonmetric distances: image retrieval and class representation,*IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 583-600, Jun. 2000.

[12] Y. Gdalyahu, D. Weinshall, and M. Warman, Self-organization in vi-sion: stochastic clustering for image segmentation, perceptual grouping, and image database organization, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1053-1074, Oct. 2001.

[13] J. Shi, and J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[14] Y. Weiss, Segmentation using eigenvectors: a unifying view, *Proc. Int. Conf. Computer Vision*, pp. 975-982, 1999.

[15] J. Costeira, and T. Kanade, A multibody factorization method for motion analysis, *Proc. Int. Conf. Computer Vision*, pp. 1071-1076, 1995.

[16] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and algorithm. *Proceedings Of Neural Information Processing Systems (NIPS)*, 2001.

[17] H. B. Luo, S. J. Zhou, Image Retrieval of Poisonous Mushrooms Based on Relevance Feedback and Clustering Algorithm, *Proceedings of the Second Intefrnational Conference on Mechatronics and Automatic Control*, pp. 685-694.

[18] J. H. Su, W. J. Huang, and V. S. Tseng, Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns, *IEEE Transactions on Knowledge and Data Engineering* vol. 23, no. 3, pp. 360-372, 2011.

[19] Z. Xiao, X. Qi, Complementary relevance feedback-based content-based image retrieval. *Multimedia Tools Appl.*, vol. 73, no. 3, pp. 2157-2177, 2014.

[20] L. Zhang, H. P. H. Shum, and L. Shao, Discriminative Semantic Subspace Analysis for Relevance Feedback, *IEEE Trans. Image Processing*, vol. 25, no. 3, pp. 1275-1287, Mar. 2016.

[21] P. F. Felzenszwalb, D. P. Huttenlocher, Efficient graph-based image segmentation, *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.

[22] T. V.Ashwin, R. Gupta, S. Ghosal, Adaptable similarity search using non-relevant information, *Proceedings of the 28th VLDB Conference*, Hong Kong, China, 2002.

[23] A. B. Benitez, M. Beigi, S. F. Chang, A content-based image meta-search engine using relevance feedback. *IEEE Internet Computing*, vol. 2, no. 4, pp.59-69, 1998.

[24] I. Bartolini, P. Ciacci, F. Waas, Feedbackbypass: A new approach to interactive similarity query processing, *Proceedings of the 27th VLDB Conference*, Roma, Italy, pp. 201-210, 2001.

[25] R. Brunelli, O. Mich, Image retrieval by examples, *IEEE Transactions on Multimedia*, vol. 2, no. 3, pp.164-171, 2000.

[26] C. Carson, S. Belongie, H. Greenspan, and J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026-1038, 2002.

[27] L. Chen, M. T. Ozsu, and V. Oria, MINDEX: An efficient index structure for salient-object-based queries in video databases. *Multimedia Systems*, vol. 10, no. 1, pp. 56-71, 2004.

[28] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments, *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20-37, 2000.

[29] M. Flickner, H. Sawhney, W. Niblack, et al., Query by image and video content: The QBIC system, *IEEE Computer Magazine*, vol. 28, no. 9, pp. 23-32, 1995.

[30] T. Gevers, A. Smeulders, Content-based image retrieval: An overview. *G. Medioni and S. B. Kang, editors, Emerging Topics in Computer Vision*, Prentice Hall, 2004.

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

[32] T. Wang, Y. Rui, S. M. Hu, Optimal adaptive learning for image retrieval, *Proceedings of IEEE CVPR 2001*, Kauai, Hawaii, pp. 1140-1147, 2001.

[33] X. S. Zhou, T. S. Huang, Comparing discriminating transfor-mations and SVM for learning during multimedia retrieval, *Proceedings of the 9th ACM Multimedia Conference*, Orlando, Florida, pp. 137-146, 2001.

[34] D. Norton, D. Heath, and D. Ventura, Annotating images with emotional adjectives using features that summarize local interest points, *IEEE Transactions on Affective Computing*, Under Review, 2016.

[35] M. Ortega-Binderberger, and S. Mehrotra, Relevance feedback techniques in the MARS image retrieval systems, *Multimedia Systems*, vol. 9, no. 6, pp. 535-547, 2004.

[36] H. T. Shen, B. C. Ooi, and X. Zhou, Towards effective indexing for very large video sequence database, *Proceedings of the ACM SIGMOD Conference*, pp. 730-741, 2005.

[37] T. T. Q. Dao, H. Q. Nguyen, V. C. Phuong, and Q. T. Ngo, An Efficient Semantic Related Image Retrieval Method. *Expert Systems with Applications*, 72, pp. 30-41, 2017.

[38] , Query refinement for multimedia similarity retrieval in MARS, *Proceedings of the 7th ACM multimedia conference*, pp. 235-238, Orlando, Florida, 1999.

[39] H. Q. Nguyen, T. T. Q. Dao, V. C. Phuong Van, V. C. Nguyen, and Q. T. Ngo, An Efficient Image Retrieval Method Using Adaptive Weights, *Applied Intelligence*, vol. 48, no. 10, pp. 3807-26, 2018.