



Low-level feature image retrieval using representative images from minimum spanning tree clustering

Piyavach Khunsongkiet¹ · Jakramate Bootkrajang¹ · Churee Techawut¹

Received: 29 June 2021 / Revised: 25 April 2022 / Accepted: 22 April 2023 /
Published online: 7 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Typical content-based image retrieval systems retrieve images based on comparison of low-level features such as images color, texture, and shapes of objects in the images. Further, the image covariance descriptor (CD) and the image Patch Relational Covariance Descriptor (PRCD) can be used to summarize low-level features and the visual arrangement to improve the precision of the retrieval. Nonetheless, comparing images based on those two descriptors is computationally expensive. Therefore, this research proposes a clustering method that dynamically groups database images using the Minimum Spanning Tree Clustering algorithm (MSTC). The technique is named Representative Images from Minimum Spanning Tree Clustering (RIMSTC). In the proposed technique, only the representative images selected from each cluster are compared with the input image. Experimental results demonstrated that the proposed representative images by COV and PRCD combined with RIMSTC helps to improve the retrieval time while maintaining comparable retrieval performance to existing methods.

Keywords Image retrieval · Image clustering and merging · Representative image retrieval · Low-level visual feature · Patch arrangement vision · Region covariance matrix · Patch Relational Covariance Descriptor

1 Introduction

Content-Based Image Retrieval (CBIR) using low-level features [22, 23, 25, 26] such as color, texture and shape, or visual arrangement is typically used to retrieve images similar

Piyavach Khunsongkiet, Jakramate Bootkrajang and Churee Techawut are contributed equally to this work.

✉ Piyavach Khunsongkiet
Piyavach.k@cmu.ac.th

Jakramate Bootkrajang
churee.t@cmu.ac.th

Churee Techawut
churee.t@cmu.ac.th

¹ Department of Computer Science, Chiang Mai University, Suthep, Chiang Mai, 50200, Thailand

to the query image from a database. The general process is to compare the input image with every stored image and return the images to the user ranked by similarity. To represent images for later arrangement similarity measurement, the Covariance Descriptor (CD) technique summarizes color and texture information, while a Patch Relational Covariance Descriptor (PRCD) is employed to describe the arrangement similarity [16].

This research provides an additional image arrangement that is an extension of the CD. To elaborate more on the visual arrangement, patches of the image are generated, and the technique calculates the similarity of covariances between the patch and its neighboring patches in the vertical and horizontal directions. After a mean similarity matrix is created from both directions which results in the visual arrangement descriptor. This visual arrangement is called the Patch Relational Covariance Descriptor (PRCD).

However, the PRCD comparison based on covariance is very time-consuming and computing the similarity between two images based on the whole features set requires a significant amount of computing power and memory. To conserve features of the descriptor without dimensionality reduction (lossy data), a clustering technique is a common technique for data analysis in many fields and seems to be appropriate for handling this requirement. Some work [12] describes the advantage of clustering which ensures better speed such as perceptual quality of watermarked images with better or comparable extraction accuracy when compromised.

This technique reduces the group by selecting a centroid data cluster and unsteadily using the centroid comparison (data approximation instead of original data reconstruction comparison). To solve the problem, the Minimum Spanning Tree Clustering algorithm (MSTC) finds groups of similar images and the representative centroid image is used in the comparison. For example, consider an input image sent to the database containing 10000 image numbers. There will necessarily be 10000 comparisons. However, if the 10000 image numbers are divided into 1000 groups and the image compared is only the centroid image (representative image of the group), the number of comparisons decreases to 1000 (10 times better). This experimental sets the number of images in the group to $N=10$, which generally means 10 retrieving items for each page. However, the clustering threshold does not separate the group of images equally.

In this research, the relaxing technique is used to merge the small clusters and collapse huge clusters, which brings each group to $N=10$. This N can be set to the other values, if necessary. These are called virtual merged and virtual collapsed clusters, and the thresholding clusters have a representative centroid image used in comparison. When the graph is cut by the MSTC threshold, it is commonly separated from the group without considering the overlapping image and the threshold is not separated equal to N numbers. This work also proposed the novel technique to evaluate the overlapping images between the virtual clusters which is equal to N . This work notes that the centroid of the MST cluster is used to compare with the input image while the virtual cluster contains the retrieving images. This proposed technique is called Representative Images from Minimum Spanning Tree Clustering (RIMSTC).

Usually, the clustering involves two steps. Firstly, the image low-level features are transformed into descriptors and evaluated the distance between those descriptors. Secondly, the descriptors of two images are used to compute a similarity score, which describes the visual similarity. This research computed all pair similarity distances during the MSTC process and passed them through the Minimum Spanning Tree (MST) algorithm [18, 30]. In addition, a more detailed history of this MST problem can be found in [11]. To create clusters from the MST, the size of the desired result is set. N is used to infer the edge cut threshold

parameter. Since there is no guarantee on the size of a cluster obtained by MST, small clusters can be merged with nearby images so that they contain roughly N members. Likewise, large clusters can be disregarded. Finally, the processing time of non-clustering and the proposed clustering-based system was compared. In addition, the results from CD and PRCD with RIMSTC were compared to SIFT with bag of feature image retrieval technique to further support the validity of using those two descriptors. The rest of the paper is organized as follows. In Section 2, the similarity measurements using descriptors, MST clustering and images retrieval framework using SIFT bag of key points features clustering, are described. In Section 3, this research introduces the image clustering with covariance descriptor and PRCD technique and gives a working example. Moreover, this research explains the idea of merging and collapsing the clusters. In Section 4, the performance of the proposed method is evaluated, and Section 5 concludes the study.

2 Literature review

Nowadays, CBIR is deployed more and more. Unlike a concept-based approach, CBIR uses an image's low-level visual information to retrieve similar images containing the same visual information as the input image. Numerous techniques have been used to automatically extract low-level visual features such as image color, texture, shape, spatial layout, from an image. This study selects some visual features which represent the visual features in this experiment study.

2.1 Color and color pattern descriptor

In the image retrieval process, the image is generally summarized in the form of visual features, and often they can later be combined and processed to construct a visual feature called a descriptor. This descriptor is used to compare the similarity between the other image descriptors. In general, humans are initially attached to color and pattern. Therefore, many techniques relate to these visual features, such as color histogram similarity, SIFT, SURF, and covariance descriptor (COV).

The scale-invariant feature transform (SIFT) is used to find salient points (key points) matching between similar points of 2 images with a tolerance of image distortion and difference. It also detects object recognition with bag of features, in which the key points are identified by a difference of gaussian (DOG) [24] and [21]. SIFT is a widely used model which is good in comparison [19]. Some research [41] applied the SIFT CBIR system while speeded up robust features (SURF) is applied approximates the DOG with box filters instead of gaussian averaging the image [2].

The covariance descriptor is the most common for describing the color information, and it is selected in our research for color similarity measurement. The covariance descriptor is proposed in [36] and some of covariance descriptor illustrations are shown in [15] Some works are currently using the advantages to perform 3D facial recognition [17]. To transform a low-level d -dimensional feature vector into a descriptor, the following equation is used.

$$C_r = \frac{1}{n-1} \sum_{k=1}^n (z_k - \mu)(z_k - \mu)^T \quad (1)$$

To compare the image descriptors, the measurement of the distance is defined as shown in (2).

$$d(A, B) = \sqrt{\sum_{i=1}^n \ln^2 \lambda_i(A, B)} \tag{2}$$

where A and B are matrices representing the covariance descriptors. $\lambda_i(A, B)$ is the eigenvalue of the pair A and B , and \ln is a natural logarithm function. Equation 2 [29] explains the relation of ratio of variances by reflecting the A and B covariance matrices when \ln is a natural logarithm function.

2.2 Visual arrangement descriptor

In addition to the color and pattern features, the arrangement of the image is another interesting feature. However, this feature is not recognized by humans. Arrangement descriptors such as HOG, PHOG, LBP and PRCD [13, 20, 38]. HOG is a fast feature descriptor used in object detection. It explains the arrangement and color at the same time, which can be used to explain the gradient histogram of each patch, while PHOG is the upgraded HOG with multiple patch sizes. Some work shows the adapted use of HOG for a recognition tool such as facial recognition by extracting HOG using the unselected magnitudes of the maximum magnitude selection method, which shows better performance [27]. Some research [35] which is referred to in [9], shows the comparison between HOG and COV. COV contributes the lowest False Positive (FPs) and False Negative (FNs) results for object recognition over the hierarchical feature-distribution (HFD).

Another applied technique called the Patch Relational Covariance Descriptor (PRCD) was employed to generate the image visual arrangement by looking at the nearest neighboring patches. To create the patch arrangement descriptor, the image used in arrangement comparison firstly needs to be minimally resized as the most appropriate resized image which is further divisible into patches. From our previous experiment, the 16-pixel patch size was the most appropriate. Therefore, the image patches were translated into the covariance descriptor in 16 pixels. The covariance distance of horizontal and vertical direction was then calculated. After that horizontal and vertical descriptors were created, z direction was evaluated by these two directions. The encoding process is shown in Fig. 1.

To compare the similarity, the relative position (x, y position) of the most similar value of the patch gradient is selected, and the difference is computed by the Euclidean distance

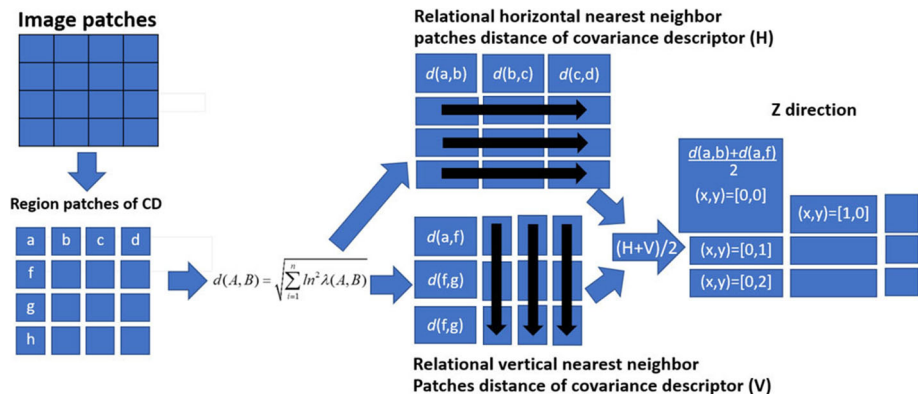


Fig. 1 The PRCD encoding process

[7]. The summation of all directions gives the arrangement similarity. A small distance value means the compared images are similar in terms of visual arrangement. In contrast, a high value means there is a difference in visual arrangement of the two images. More implementation detail, it was shown in the methodology section.

2.3 Existing image retrieval system

A common Content-Based Image Retrieval (CBIR) process is a concept to retrieving image which is similar to an input image which is referred to a descriptor or encoder, and implement a powerful CBIR has the relevance component such as creating index, clustering data, relevance feedback (RF) and etc. Recently, some research [34] shows a powerful technique combining some of the common processes above to extract different features form an image such texture, color and shape and integrate them into a hybrid feature matrix or vector (HFV), which is the input for an extreme learning machine (ELM) to learn for retrieving a relevance set of images. Moreover, this ELM is also integrated with RF to be an ELM-RF framework, which makes the CBIR a compatible system. In addition, some work [4] presents the interesting idea for image segmentation based on a supervised learning technique such using a support image (image sample baseline) to guild the deep learning network (such VGG and Resnet network model) and increase the efficient and decoding result to segmentation area (such a lesion or cancer area output). This work looks similar to the well-known autoencoder network technique [1]. Other interesting work [39] shows the technique to discriminate a good proposal region object with a single label from the image using multi-label management. After that, the discriminative regions are aggregated to obtain high-confidence seeds that are used to grow deep learning networks. Moreover, some work [5] propose a joint framework containing a transfer learning strategy and a deep super-resolution framework to generate high-resolution slice images from low-resolution ones. Next, works are proposed image segmentation technique using deep learning model such as [10] which is shown technique to build a segmentation network and [4] which firstly evaluates a global feature and discriminate general those segmentation area such a lesion area by using Global Class Activation (GCA) module. After that Local Bin Excitation (LBE) module is used to extract excited lesion features in a local manner and allows the lesion regions to be more fine-grained. Another work [3] is a mixed feature (such as fundus image, visual field tests and age) used to train the network multi-modality fusion learning. Some CBIR research shows a combination of three different approaches such as local mesh peak valley edge pattern (LMePVEP), local mesh ternary pattern (LMeTerP) and texture gradient-based images. It proposes a feature vector which gives the better accuracy of image retrieval.

In this work, this study is interested in a clustering technique and combine with low-level features image retrieval. This conducts technique based on unsupervised learning and plugs to some of related systems as a microservice idea [28].

2.4 Data clustering

In general, it is difficult to search for similar images due to the large size of image descriptors. Restricting the comparison to the representative descriptors is expected to reduce the search time. In addition, clustering in image retrieval is a method to scalable and proposed faster and more efficient. Developers realized that standalone computers are necessary to be distributed to serve clients by clustering techniques for cloud systems [12]. This can be combined with many recent image retrieval techniques such as encoding to index data, without indexing and otherwise.

To cluster data, common techniques are MSTC, K-mean, K-medoid, Agglomerative clustering, DBscan, etc [14, 32, 37]. In this work, the MSTC technique is commonly preferable because this technique does not need to specify the number of clusters and the threshold variable can be changed over a setting up heuristic algorithm (HA). The edge cut threshold parameter is instead used to evaluate the cluster. However, the common threshold clustering technique cannot identify the number of data in each cluster. Moreover, it is hard to evaluate the overlapping data. The MSTC technique is, therefore, applied to find groups of similar images in this research. The representative images for the group of similar images were selected the similarity was calculated with the descriptor of the query image to find out the most similar group. Additionally, the closest similar images could be merged to meet the required result set size. Some related work shows the techniques to select a color threshold based on HA such as [31]. This work shows MFO helps to get a satisfactory result.

2.5 Image retrieval using bag of features clustering with scale

Invariance Feature Transform SIFT Algorithm [24] is used to select the salient points and match the similarity of the image. In addition, it can resolve the images which differ in resolutions and rotations.

Some related work using SIFT with Bag of features is shown in [6]. To cluster similar images using SIFT, the images key points are firstly extracted as the descriptors. After that, every image's key points in the same class are pooled together and clustered by K-mean. The clusters of key points from every image are converted to the cluster bins of histogram. To compare the input image with the database images, every key points extracted from the input image is compared with the closest centroid value to select the cluster of each key point. Bins of input image histogram are then created and compared with the bins of images in the database. Finally, the top-n most similar images in the database are retrieved for the user. In our research, this method is the baseline algorithm for comparing our MSTC with PRCD technique. The sift clustering is shown in Fig. 2.

In addition, SIFT is widely adapted to create the image retrieval system. Some work [40] has shown the complementary of SIFT and deep learning to adjust the searching of relevant object image. Therefore, SIFT is selected for our baseline to compare with our technique in this work while the other methods are combined with our representative clustering technique.

3 Methodology

3.1 Proposed framework

To create a low-level image clustering retrieval system, this research proposes 2 main steps. In the first step (database clustering preparation), the images in the database were clustered and the centroid representative image was selected for use in measuring the similarity between the input image and images in the database. However, Nearest neighbor images could be merged to the clusters in case of inadequate images in a cluster in order to meet the required result set size. The clusters can also be collapsed in case of many retrieved images. The created clusters from merged images and collapsed clusters are called the visual merged and collapsed clusters. In the second step (data retrieval step), the input image was compared with the representative image of each cluster to retrieve the most similar image group from the visual merged and collapsed clusters.

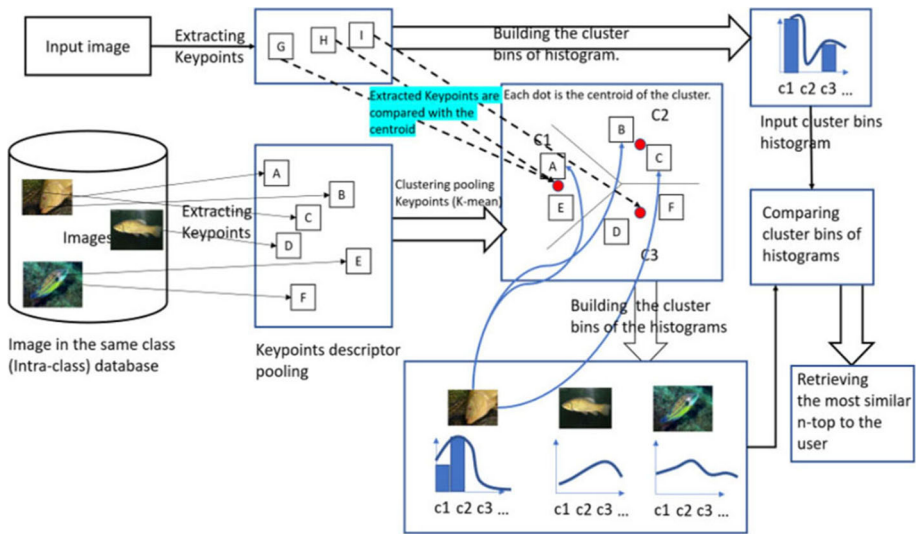


Fig. 2 SIFT Image retrieval conceptual framework

3.2 Step 1: preprocessing clustering images in database

Step 1 is shown as in Fig. 3.

In Fig. 3, the visual color and arrangement features of the images are translated to CD and PRCD, respectively. Next, the distance matrix for each descriptor was created. The process for creating the distance matrix for each descriptor, is shown in Fig. 4.

In Fig. 4, the features of all images are encoded to covariance descriptors. All pairs of the descriptors were used to compute the distance to obtain the distance covariance descriptor pair matrix.

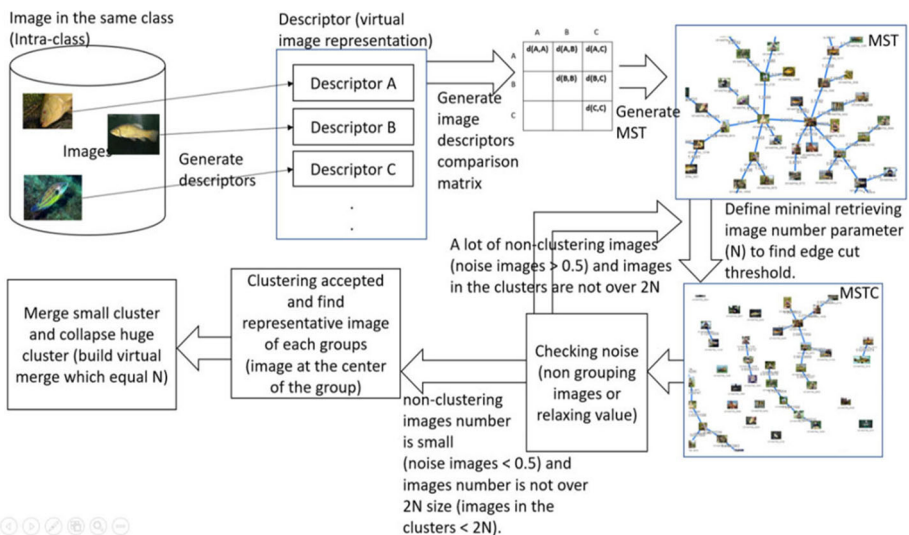


Fig. 3 Step 1: The image clustering preparation of RIMSTC framework

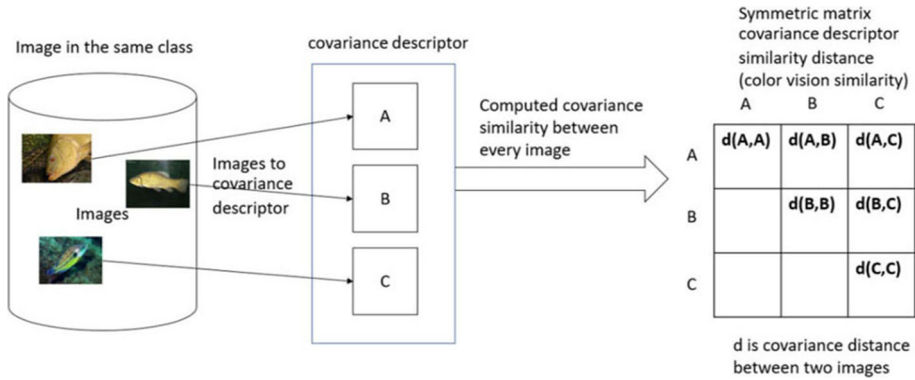


Fig. 4 The translation of covariance descriptor to distance matrix

In Fig. 5, the images patches were encoded into the PRCD.

To generate PRCD, an image firstly needs to resize at the minimum which can be zero-modulation by the patch size ($s \times s$).

Algorithm 1: Generate patches matrix from a minimum resize image.

After the matrix P of s size patch image is evaluated, the covariance distance of the nearest neighbor descriptor is calculated and determined on horizontal and vertical axis. The pseudocode to generate PRCD horizontally is shown in Algorithm 2. To encode the vertical PRCD direction, the similar code is simulated from Algorithm 2 or the transpose matrix on the input image with Algorithm 2.

From Algorithm 2, f is a function to return image feature when cov function is generated covariance descriptor. $Dcov$ is a function to measure similarity between 2 covariance descriptors.

To generate the vertical PRCD direction, the similar code in Algorithm 2 can be repeatedly processed or initially use the transpose matrix on the input image. In addition, the average gray intensity of each patch ($S \times S$) of the image needs to be integrated to the PRCD which is used for boosting the patch arrangement PRCD flow of different objects of the same label as in Algorithm 3. Notice that, a color (RGB) did not used in this step. For example, the example color channel $R=50$ is a different color vision with $G=50$ while the gray

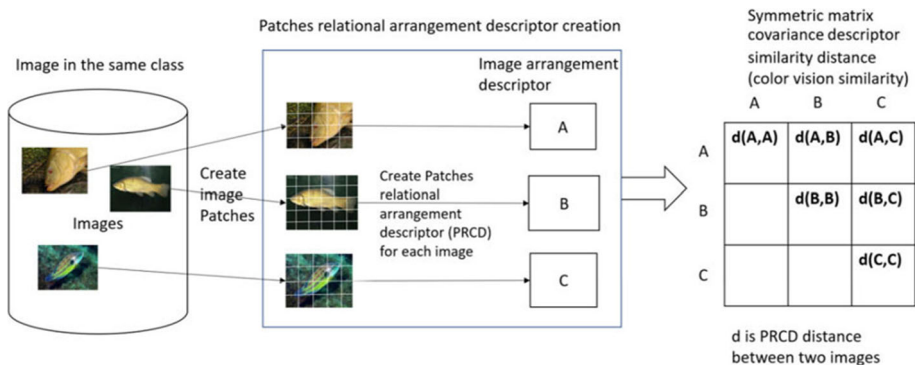


Fig. 5 The translation of PRCD to distance matrix

Require: I is an image, s is the size of the patch

```

1:  $w, h \leftarrow f_{\text{shape}}(I)$ 
2:  $w_{\text{num}} \leftarrow \lfloor \frac{w}{s} \rfloor + 1, h_{\text{num}} \leftarrow \lfloor \frac{h}{s} \rfloor + 1$ 
3:  $w_{\text{new}} \leftarrow w_{\text{num}} \times s, h_{\text{new}} \leftarrow h_{\text{num}} \times s$ 
4:  $w_{\text{new}}, h_{\text{new}} \leftarrow \text{minimum resize on } (I, s)$ 
5:  $I_{\text{new}} \leftarrow f_{\text{resize}}(I, w_{\text{new}}, h_{\text{new}})$ 
6:  $P \leftarrow []$ 
7: for  $i \leftarrow 1$  to  $w_{\text{num}}$  do
8:    $\text{row} \leftarrow []$ 
9:   for  $i \leftarrow 1$  to  $w_{\text{num}}$  do
10:      $p$  is function to generate patch
11:      $//$  windows at position  $i, j$  by  $s$ 
12:      $p(I_{\text{new}}, i, j)$  pushes to row
13:   end for
14:   row pushes to  $P$ 
15: end for
16: return  $P$ 

```

Algorithm 1 Calculate patching image with minimum resize.

intensity is the same. This is similar to HOG. However, this work used the average patch gray intensity.

To evaluate PRCD similarity, this process shown in Algorithm 5.

To build the MST, pairs of the descriptors were then used to compute the distance PRCD descriptor pair matrix. Next, the distances matrix was evaluated to form an MST.

The objective is that clusters do not contain the image more than N ($N=10$). Following this constraint, a list of thresholds from sub-MST needs to be evaluated from the MST. For each sub-MST, a maximum value of the edge list of sub-MST is the minimum candidate threshold that uses for candidates to other clusters or sub-MSTs. The minimum threshold

```

1:  $H \leftarrow []$ 
2:  $v \leftarrow \text{len}(P), h \leftarrow \text{len}(P[0])$ 
3: for  $j \leftarrow 1$  to  $v$  do
4:    $H_{\text{row}} = []$ 
5:   for  $i \leftarrow 1$  to  $h$  do
6:     if  $i + 1$  not more than  $h$  then
7:        $C_1 \leftarrow \text{cov}(f(P(I_{\text{new}}, i, j)))$ 
8:        $C_2 \leftarrow \text{cov}(f(P(I_{\text{new}}, i + 1, j)))$ 
9:        $d \leftarrow D_{\text{cov}}(C_1, C_2)$ 
10:       $d$  pushes to  $H_{\text{row}}$ 
11:     end if
12:   end for
13:    $H$  pushes to  $H_{\text{row}}$ 
14: end for
15: return  $H$ 

```

Algorithm 2 H horizontal PRCD evaluation.

Require: I_patch is an image, w_{num}, h_{num} is the patch size

```

1: gray_avg = []
2: for i ← 0 to w_num - 1 do
3:   gray_average_row ← []
4:   for j ← 0 to h_num - 1 do
5:     if i + 1 < w_num + 1 and j + 1 < h_num + 1 then
6:       //gray_average is to find average gray patch intensity
7:       gray_average_row.append(gray_average(I_patch[i][j]))
8:     end if
9:   end for
10:  gray_average.append(gray_average_row)
11: end for
12: return gray_avg

```

Algorithm 3 Calculate gray intensity.

(t_{min}) needs to evaluate from minimum candidate thresholds of every threshold of all sub-MST. Each sub-MST equation is evaluated as follows.

$$\arg \min_D \left(\sum_{j=1}^{n-1} d(I_i, J_j) \right) \text{ when } D \in d_1, \dots, d_{n+1} \quad (3)$$

The list of minimum distances of image I_i is $d(I_i, J_j)$ when d is a similarity function between 2 images and D_i is defined to contain the list of the i sub-MST. Note that the i of D equals to $n + 1$ because each cluster has member numbers equal to N .

$$d_{max}(i) = \max(D_i) \quad (4)$$

$$D_{max} = d_{max}(1), d_{max}(2), \dots, d_{max}(v) \quad (5)$$

From above formula, v is number clusters that contain N images. $d_{max}(i)$ is the max distance between two images from i sub-MST position and D_{max} contains candidate d_{max} from all subMSTs.

Require: $gray_avg$ is matrix of gray average patch intensity

Require: w_{num}, h_{num} is the patch size

```

1: prcd_new ← [], max_sim ← 50//maximum COV similarity
2: for i ← 0 to w_num - 1 do
3:   prcd_row ← []
4:   for j ← 0 to h_num - 1 do
5:     value ← PRCd[i][j], intensity ← gray_avg[i][j]
6:     prcd_row.append(((max_sim - value)/max_sim) × intensity)
7:   end for
8:   prcd_new.append(prcd_row)
9: end for
10: return PRCd_new

```

Algorithm 4 Integrated gray intensity with PRCd.

Require: $PRCD1, PRC D2$

```

1:  $d\_sum, count \leftarrow 0.0$ 
2:  $(x1, y1) \leftarrow shape(PRC D1), (x2, y2) \leftarrow shape(PRC D2)$ 
3: for  $i \leftarrow 0$  to  $x1$  do
4:   for  $j \leftarrow 0$  to  $y1$  do
5:      $k\_min, l\_min \leftarrow -1$ 
6:      $min\_d, min\_d\_eud \leftarrow infinity$ 
7:     for  $k \leftarrow 0$  to  $x2$  do
8:       for  $l \leftarrow 0$  to  $y2$  do
9:          $d \leftarrow abs(PRC D1[i][j] - PRC D2[k][l])$ 
10:        if  $min\_d > d$  then
11:           $min\_d \leftarrow d, k\_min \leftarrow k, l\_min \leftarrow l$ 
12:        end if
13:      end for
14:    end for
15:     $d\_sum \leftarrow d\_sum + \sqrt{(i - k\_min)^2 + ((j - l\_min)^2)}$ 
16:     $count \leftarrow count + 1.0$ 
17:  end for
18: end for
19: return  $d\_sum/count$ 

```

Algorithm 5 Calculate PRCD similarity distance.

To select the threshold that makes all clusters do not contain the image more than N ($N=10$), $d_{max}(i)$ has to be the lowest value.

$$t_min = \min(D_{max}) \quad (6)$$

This threshold t_min is used for the thresholding edge for the MSTC.

However, some selected thresholds generated small number clusters which has a lot of porous image space. in Fig. 6, the minimum selected threshold from the candidate n -top ($t_min(n = 5)$) was not appropriated for clustering for the following.

Therefore, thresholds more than $t_min(n = 10)$, were selected to solve this problem. This is called the “relaxation process”. In the relaxation process, an accepted value (a) was defined and calculated from the number of connected images divided by the total number of images as shown in the following equation.

$$a = \frac{\text{number of connected edge}}{\text{number of total edge}} \quad (7)$$

This a is called the “relaxation parameter”.

In this case, $a = 1$ means all nodes were connected to at least one and $a = 0$ means no connected nodes. In our research, a was set at 0.5. If the candidate t_min produced an a less than 0.5, the next minimum t_min was continuously selected until a was more than 0.5.

In addition, the maximum bound condition (b) is fixed for the number of images in each cluster to be not more than $2n$ (Because its $n = 10$, the maximum of each cluster should be not more than 20). Thus, b could be set between n and $2n$. The algorithm is follows.

In Fig. 7, the new cluster C1 on the right appeared. After the threshold was adjusted, C1 and C2 on the left were integrated to C2 on the right and porous image space decreased. Figure 8 is the zooming result of CD with RIMST.

Figures 9 and 10 show an example result of PRCD with relaxation MSTC process.

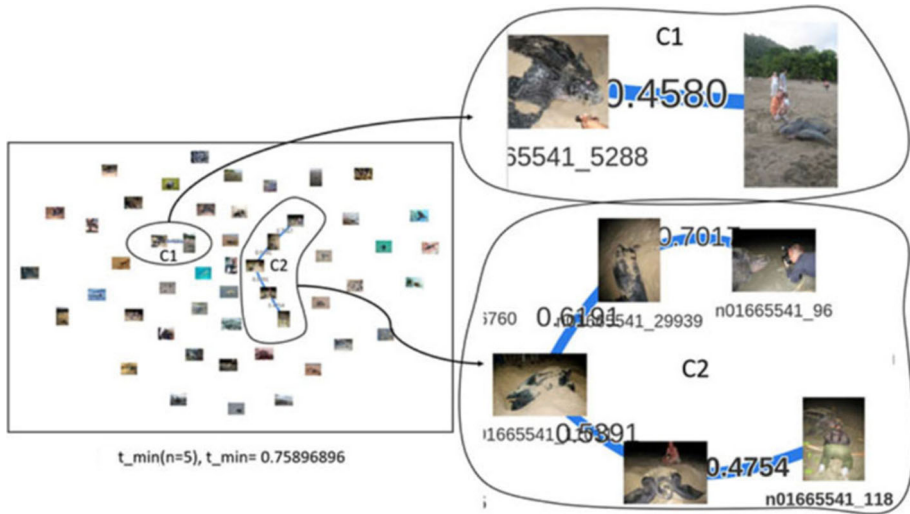


Fig. 6 Example of Non-clustering images with inappropriate threshold ($n = 5$) of CD

In addition, PRCD Cluster C2 (of Fig. 10) shows that the background color is different when bird image arrangements look similar. Both images in C2 have the same value of the average edge distance. This means any of them can be a centroid image (Table 1).

Sometimes the relaxation parameter (a) of (7) is closed to 0.5 and the data is scattered or is not satisfied (feeling many data are no group or non-cluster data). In this case, those non-clustering can abandon in considering to retrieve for the result set of image retrieval.

Require: *gray_avg* is matrix of gray average patch intensity

```

1:  $N \leftarrow 10$ 
2:  $n \leftarrow 10$ 
3:  $a \leftarrow 0$ 
4: //  $T_{relaxing\_parameter}$  set to half porous
5:  $T_{relaxing\_parameter} = 0.5$ 
6: while  $try\_again == true$  do
7:   // Relaxation constraints
8:   if  $a < T_{relaxing\_parameter}$  and  $n < 2 \times N$  then
9:      $t\_min \leftarrow findt\_min$  from  $n$ 
10:     $images\_in\_cluster \leftarrow get\ number\ of\ images\ in\ cluster\ from\ MSTC$ 
11:    // check relaxation parameter
12:     $a \leftarrow \frac{images\_in\_cluster}{total\_images}$ 
13:     $try\_again = true$ 
14:     $n \leftarrow n + 1$ 
15:  else
16:     $return\ t\_min, n$ 
17:  end if
18: end while

```

Algorithm 6 Relaxation process and find t_min and relaxing n .

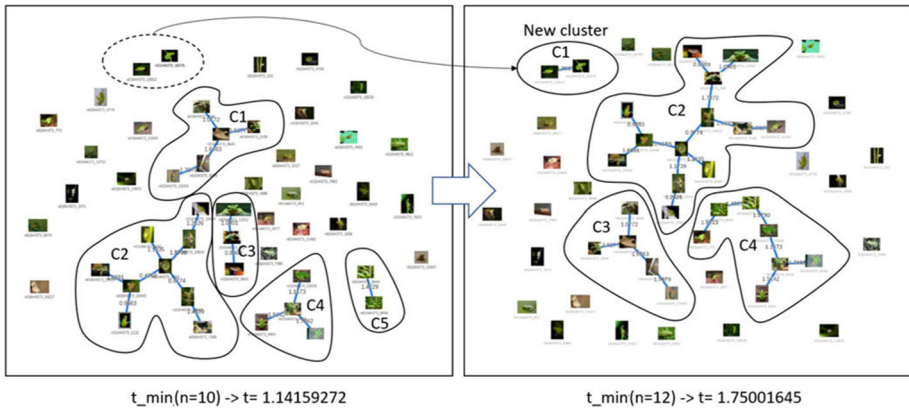


Fig. 7 The image clustering is evaluated over the relaxation processing and the new threshold is selected to compact clusters of CD with RIMSTC

Otherwise, those non-clustering data can repeatedly run to the MSTC and relax process (RIMSTC) again when user needs more clusters with lower porous, lower density.

3.3 Step 2: input image comparison and retrieving the most similar cluster

However, the number of members in some clusters has both lower than N and more than N . This can occur because the MSTC edge cut threshold does not guarantee the number of the images in each cluster. This problem was fixed by the visual merging and collapsing cluster process. To merge the nearest images to insufficient clusters, the nearest image of sub-MST is included in the small clusters until it reaches N . Some overlap might occur.

Figure 11 shows that C2 expanded to $N=10$ and was generated by sub-MST. This new cluster is called a virtual merging cluster.

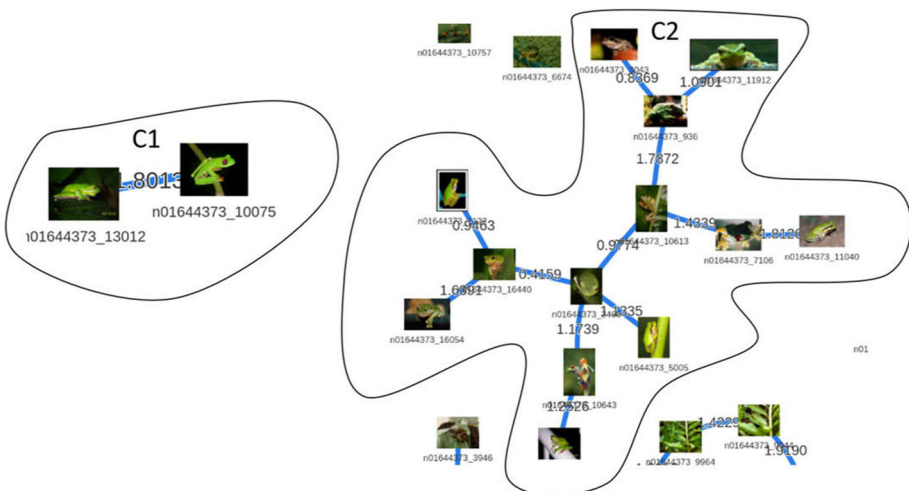


Fig. 8 Some zooming example results of new cluster C1 and C2

PRCD $t_{min}(N=19) \rightarrow t=6.68295$,
relaxing value $a=0.40816$

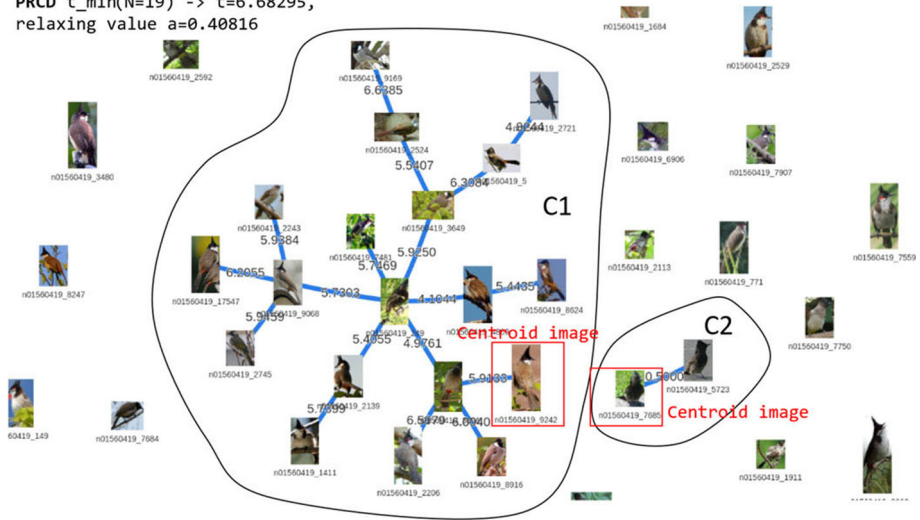


Fig. 9 Some example results of PRCD with relaxation MSTC process

Figure 12 shows that C1 collapsed to $N=10$. This cluster is called a virtual collapsing cluster. In addition, some images overlapped between C2 virtual merging cluster and C1 virtual collapsing cluster (Fig. 13).

The conclusion of the overall framework is shown in Fig. 14.

Examples of retrieved images are shown in Fig. 15 with the query image as a picture of a bird in the first row.

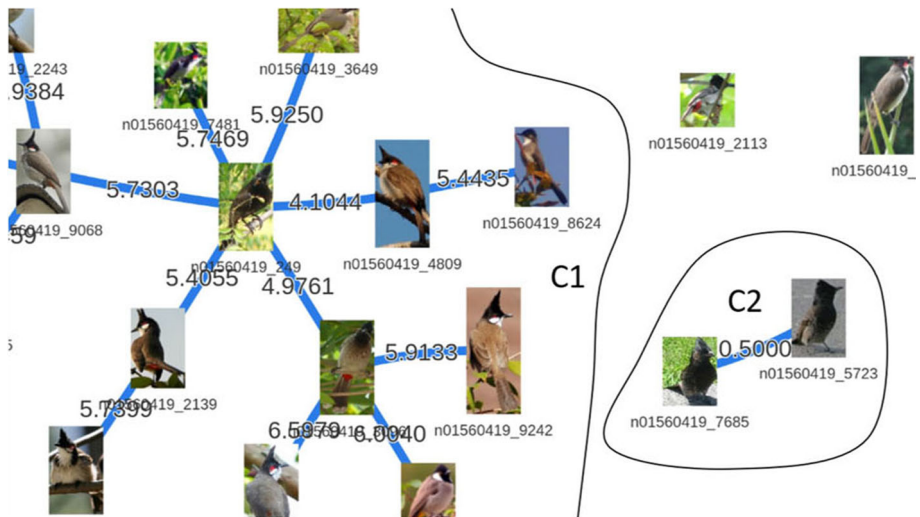


Fig. 10 Some zooming example results of PRCD with MSTC of Fig. 9. In C2 cluster, bird image arrangements is similar

Table 1 Example of PRCD with MSTC and centroid image label (n01560419)

Cluster no	Centroid	Image (average distance of the image node)
C1	n01560419_9169	n01560419_8916 (6.0040), n01560419_3096 (5.8728), n01560419_2206 (6.5979), n01560419_9242 (5.9132), n01560419_249 (5.3146), n01560419_2139 (5.5726), n01560419_4809 (4.7739), n01560419_7481 (5.7468), n01560419_9068 (5.9550), n01560419_3649 (5.9247), n01560419_1411 (5.7399), n01560419_8624 (5.4435), n01560419_17547 (6.2055), n01560419_2745 (5.9458), n01560419_2243 (5.9384), n01560419_5 (5.6164), n01560419_2524 (6.0896), n01560419_2721 (4.9243), n01560419_9169 (6.6385)
C2	n01560419_5723	n01560419_7685 (0.5), n01560419_5723 (0.5)

The top image is the input image. The result A is from PRCD, the result B is from SIFT, and the result C is from the CD.

4 Experimental results

This research objective is to evaluate the result in two parts. Firstly, this work compared the effectiveness of CD + RIMSTC (color vision), PRCD + RIMSTC (image arrangement vision), and SIFT with K-mean clustering technique. Secondly, this research compared the effectiveness of the execution time between using cluster-based techniques and non-clustering retrieval. To evaluate the image retrieving similarity performance, 10 classes from the ImageNet training dataset were selected [33], each class containing 50 images with RGB channels (Not grayscale image). For each class, the images were cropped to image objects

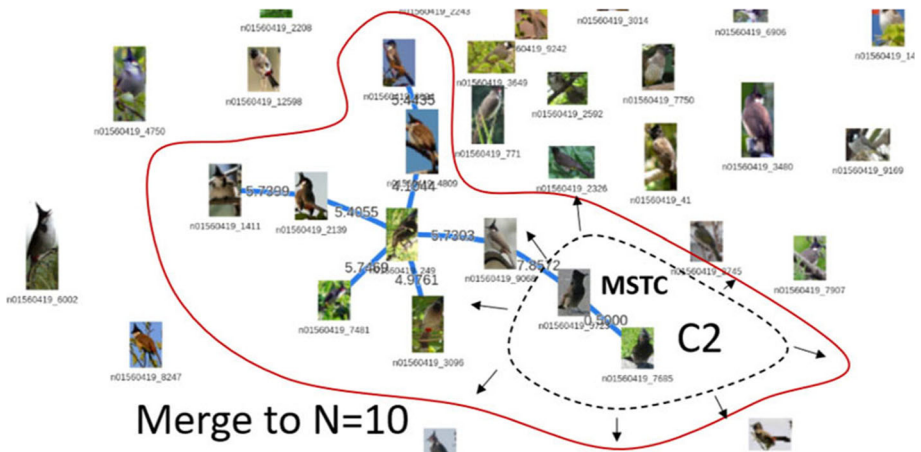


Fig. 11 C2 is expanded to N=10 to be virtual merged cluster

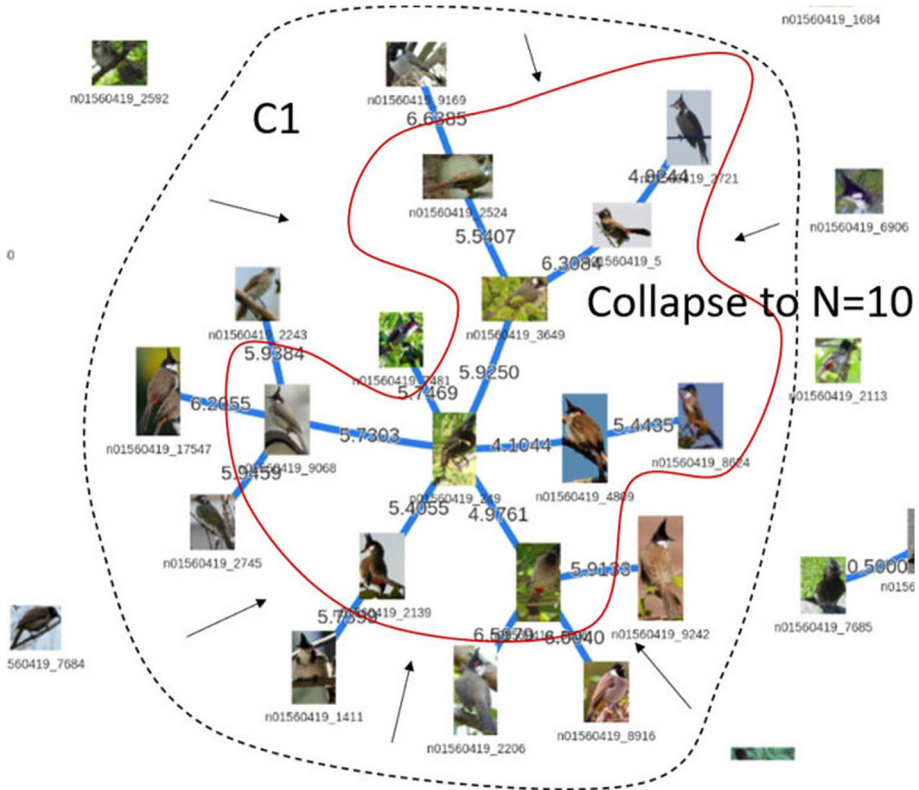


Fig. 12 C1 is collapsed to N=10 to be virtual collapsed cluster

and preprocessed following our RIMSTC framework and SIFT clustering, and the 10 validating images from each class were selected randomly and the retrieving result set was computed. Finally, this input image and result set of images without cropped object were given to 48 users to select the most similar retrieving result set based on their own judgment. These results are taken as ground truths. To evaluate the effectiveness of each method, Average Precision (*AP*) [8] was computed with in the following equation.

$$AP_j = \sum_{k=1}^n (p(C_k)) \tag{8}$$



Fig. 13 examples of overlapping image between C2 virtual merging cluster and C1 virtual collapsing cluster

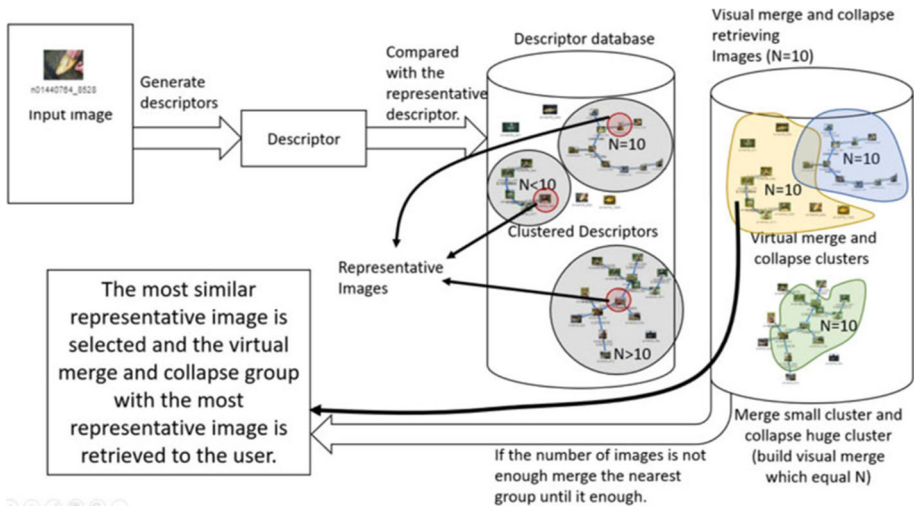


Fig. 14 The conceptual of RIMSTC framework of input image comparison and retrieving the most similar cluster

where AP_j of class j and p is the precision of each sample and m_j is the total number of the sample. C is the set of the retrieving image. To find the mean average precision, AP of each method is added together and divided by the total number of $AP(n)$.

$$MAP = \sum_{i=1}^{m_j} \left(\frac{AP}{n} \right) \tag{9}$$

SIFT with K-mean clustering shows the best performance for selecting the similar scaling and rotating images, and for reducing the number of considering features. However, some images produce a small number of key points. This makes the bin was created from the key points too small for comparison. CD and PRCD are compared with SIFT and consider the color and arrangement vision respectively. Table 2 shows that every visual perception is important to consider as they have similar MAP values. To compare the performance between non-clustering and clustering, the comparison following each technique was separated by drawing the graph of the execution time of CD and PRCD respectively. After that the 50 images were sent to the system the execution time from the descriptor of the input image was captured. This simulates the execution time of the server side, and the descriptors of every image are extracted, ready for the comparing steps. In Fig. 16, the top line is the CD non-clustering, and the bottom line is the CD with the proposed clustering technique which uses the representative images during comparison. In Fig. 17, the top line is the PRCD non-clustering, and the bottom line is PRCD clustering, and it uses the representative image instead. It shows that CD and PRCD with clustering techniques can help decrease the execution time for retrieving the image result sets. Moreover, it shows a reduction in the frustrating value of the execution time.

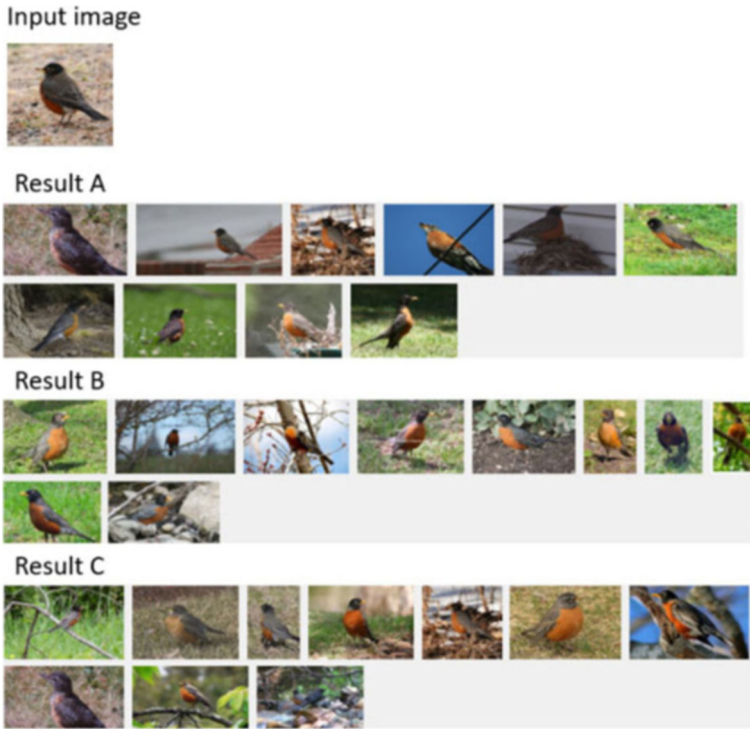


Fig. 15 The example of the similar image retrieval result

5 Discussions and concluding remarks

This work proposed the technique RIMSTC to cluster $N=10$ images while the common MSTC does not equally separate the $N = 10$ image number by using the threshold. Moreover, RIMSTC shows how to merge the small cluster and collapse the huge cluster (virtual merged and virtual collapsed clusters) while it is still sufficient to $N=10$. In addition, the thresholding clusters (MSTC) is used in comparison with an input image. Our RIMSTC also shown some images can be in multiple clusters and this makes those overlapping images can be retrieved more than once by our threshold conditional and relaxation parameter. The overlap images do not need to be in the single cluster. This work also proposed to decrease the execution time by using RIMSTC with representative images to retrieve similar groups of images instead of comparing the input image with every image in the database and return the n-top similar results to the user. This work uses the technique of creating a virtual cluster by merging and collapsing clusters less than $N=10$. The representative image used is the centroid image which appears in the MST clusters and compare with the input image while the virtual cluster (contains N images) is used for retrieving N images as in Fig. 12.

Table 2 Mean Average Precision result of 4800 sample

SIf K-mean clustering	CD+RIMSTC	PRCD+RIMSTC
0.4910	0.4919	0.5173



Fig. 16 Execution time of CD with non-clustering and RIMSTC for 50 times image retrieving

In conclusion, this RIMSTC has 2 main properties such as clustering based density of data, relaxing density for containing N data.

This work also shows the MAP result comparison between descriptors that were integrated with RIMSTC technique and SIFT technique. Table 2 shows the integrated descriptors with RIMSTC, and our clustering technique is slightly different MAP result when compared to the SIFT technique. It means that users were attracted by variety similarity visions such as color and image arrangement vision. In addition, multiple visions and SIFT could be combined and considered together and used in measuring image similarity.

To apply other techniques, this work notes that the descriptors or image encoder can be changed to the other methods and adapted to this representative clustering technique. To decrease execution time between the non-clustering and clustering method, this research firstly proposed the technique to select the representative image by using our MSTC and sought to relax the edge cut threshold and select the nearest centroid image, which is the representative image to compare with input image. Moreover, this work showed the technique to merge clusters with an insufficient number of images and collapse clusters with an excessive number of images down to 10.

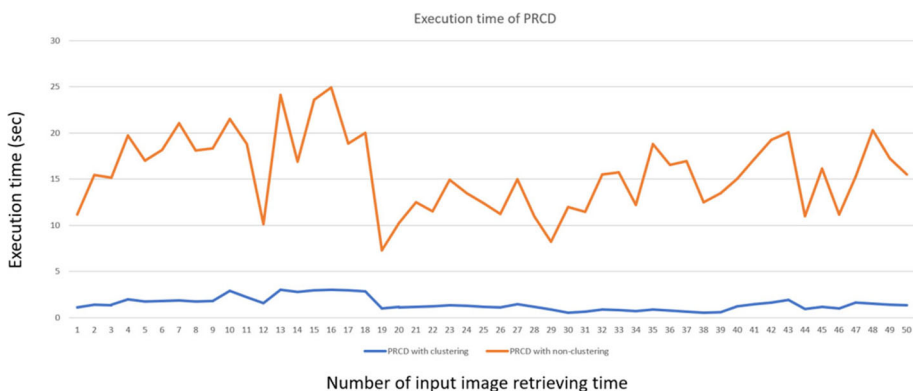


Fig. 17 Execution time of PRCD with non-clustering and RIMSTC for 50 times image retrieving

The graphs in Figures 16 and 17 show that integrated MSTC can help to decrease the execution time for retrieving image result sets. Moreover, it reduces the graph frustration.

However, this clustering technique (RIMSTC) helps faster in retrieving images. It trades off the execution times in clustering processes such as evaluating the distance of the fully connected graph before creating the MST and creating descriptors. This needs to be realized before using this technique. For the long term CBIR, huge system or without compacting image, this technique should be applied and take advantage of scalable and faster image retrieval advantages. Another problem is that PRCD is unsupervised technique which cannot considered a background image object and the background noise are translated by PRCD. To solve this problem, powerful deep learning such a masked object segmentation encoder or decoder deep learning can be applied with PRCD in further work.

For future work, this work aims to combine high-level semantic meaning [22, 23, 25, 26] and low-level similarity visions to reduce the gap off human similarity perception.

Funding The authors declare that no grants supported.

Declarations

Competing interests Non-financial interests: No funding was received for conducting this study.

References

1. Bank D, Koenigstein N, Giryes R (2020) Autoencoders. Machine Learning (cs.LG); Computer Vision and Pattern Recognition (cs.CV); Machine Learning (stat.ML)
2. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. *Comput Vis-ECCV 2006* 3951:404–417. https://doi.org/10.1007/11744023_32
3. Cao Z et al (2021) Multi-modality fusion learning for the automatic diagnosis of optic neuropathy. *Pattern Recogn Letters* 142:58–64. <https://www.sciencedirect.com/science/article/pii/S0167865520304402>. <https://doi.org/10.1016/j.patrec.2020.12.009>
4. Chen T et al (2021) Discriminative cervical lesion detection in colposcopic images with global class activation and local bin excitation. *IEEE J Biomed Health Inform PP*
5. Chen J et al (2021) A transfer learning based super-resolution microscopy for biopsy slice images: the joint methods perspective. *IEEE/ACM Trans Comput Biol Bioinforma* 18(1):103–113. <https://doi.org/10.1109/TCBB.2020.2991173>
6. Deselaers T, Pimenidis L, Ney H (2008) Bag-of-visual-words models for adult image classification and filtering
7. Dokmanic I, Parhizkar R, Ranieri J, Vetterli M (2015) Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Proc Mag* 32(6):12–30. <http://arxiv.org/abs/10.1109/MSP.2015.2398954>
8. Dupret G, Piwowarski B (2013) Model based comparison of discounted cumulative gain and average precision. *J Discrete Algorithm* 18:49–62. <https://doi.org/10.1016/j.jda.2012.10.002>, selected papers from the 18th International Symposium on String Processing and Information Retrieval (SPIRE 2011)
9. Faulkner H et al (2015) A study of the region covariance descriptor. Impact of feature selection and image transformations
10. Feng R et al (2021) A deep learning approach for colonoscopy pathology WSI analysis: accurate segmentation and classification. *IEEE J Biomed Health Inform* 25(10):3700–3708
11. Graham R, Hell P (1985) On the history of the minimum spanning tree problem. *Ann Hist Comput* 7(1):43–57. <https://doi.org/10.1109/MAHC.1985.10011>
12. J A, R S (2022) A faster secure content-based image retrieval using clustering for cloud. *Expert Syst Appl* 189:116070. <https://www.sciencedirect.com/science/article/pii/S0957417421014093>. <https://doi.org/10.1016/j.eswa.2021.116070>
13. Jamshed M, Parvin S, Akter S (2015) Significant hog-histogram of oriented gradient feature selection for human detection. *Int J Comput Appl* 132:20–24. <https://doi.org/10.5120/ijca2015907704>
14. Kanagala H, Krishnaiah V (2016) A comparative study of k-means dbscan and optics

15. Karacan L, Erdem E, Erdem A (2013) Structure-preserving image smoothing via region covariances. *ACM Trans Graph* 32(6). <https://doi.org/10.1145/2508363.2508403>
16. Khunsongkiet P, Bootkrajang J, Techawut C (2020) Patch relational covariance distance similarity approach for image ranking in content-based image retrieval
17. Križaj J, Dobrišek S, Štruc V (2022) Making the most of single sensor information: a novel fusion approach for 3d face recognition using region covariance descriptors and gaussian mixture models. *Sensors* 22:2388. <https://doi.org/10.3390/s22062388>
18. Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc Am Math Soc* 7(1):48–50. <https://doi.org/10.2307/2033241>, full publication date: Feb. 1956
19. kumar Panda S, Panda CS (2019) A review on image classification using bag of features approach. *Int J Comput Sci Eng* 7:538–542. <https://doi.org/10.26438/ijcse/v7i6.538542>
20. Liao S, Law MWK, Chung ACS (2009) Dominant local binary patterns for texture classification. *IEEE Trans Image Process* 18(5):1107–1118. <https://doi.org/10.1109/TIP.2009.2015682>
21. Lindeberg T (2012) Scale invariant feature transform. *Scholarpedia* 7:10491. <http://arxiv.org/abs/10.4249/scholarpedia.10491>
22. Liu Y, Zhang D, Lu G, Ma W-Y (2007) A survey of content-based image retrieval with high-level semantics. *Pattern Recogn* 40(1):262–282. <https://www.sciencedirect.com/science/article/pii/S0031320306002184>. <https://doi.org/10.1016/j.patcog.2006.04.045>
23. Liu Y, Zhang D, Lu G, Ma WY (2007) A survey of content-based image retrieval with high-level semantics. *Pattern Recogn* 40(1):262–282. <https://doi.org/10.1016/j.patcog.2006.04.045>
24. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60:91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
25. Lu Y, Zhang L, Tian Q, Ma W-Y (2008) What are the high-level concepts with small semantic gaps?
26. Memar Kouchebagh S, Suriani Affendey L, Mustapha N, Doraisamy SC, Ektefa M, Lukose D, Ahmad AR, Suliman A (2012) (eds) High level semantic concept retrieval using a hybrid similarity method. In: Lukose D, Ahmad AR, Suliman A (eds) *Knowledge technology*. Springer, Berlin, pp 262–271
27. Nguyen-Quoc H, Hoang VT (2021) A revisit histogram of oriented descriptor for facial color image classification based on fusion of color information. *J Sens* 2021:6296505. <https://doi.org/10.1155/2021/6296505>
28. Pachghare V (2016) Microservices architecture for cloud computing. *J Inf Technol Sci* 2:13
29. Peters G, Wilkinson JH (1970) $Ax = \lambda bx$ and the generalized eigenproblem. *SIAM J Numeric Anal* 7(4):479–492. <https://doi.org/10.1137/0707039>
30. Prim RC (1957) Shortest connection networks and some generalizations. *Bell Syst Technic J* 36(6):1389–1401. <https://doi.org/10.1002/j.1538-7305.1957.tb01515.x>
31. Rajinikanth V, Kadry S, Crespo RG, Verdú E (2021) A study on rgb image multi-thresholding using kapur/sallis entropy and moth-flame algorithm. *Int J Interact Multimed Artif Intell* 7:163–171. <https://doi.org/10.9781/ijimai.2021.11.008>
32. Rakesh C, Sarma C, Jha M (2015) Document clustering using k-means and k-medoids
33. Russakovsky O et al (2015) Imagenet large scale visual recognition challenge. [arXiv:1409.0575](https://arxiv.org/abs/1409.0575)
34. Shikha B, Pandove G, Dahiya P (2020) An extreme learning machine-relevance feedback framework for enhancing the accuracy of a hybrid image retrieval system. *Int J Interact Multimed Artif Intell* 6:15–27. <https://doi.org/10.9781/ijimai.2020.01.002>
35. Sulic VS, Perš J, Kristan M, Kovacic S (2010) Histogram of oriented gradients and region covariance descriptor in hierarchical feature-distribution scheme. In: *Proceedings of the 19th international electrotechnical and computer science conference (ERK2010) vol 2010*. <http://vision.fe.uni-lj.si/docs/danas/SulicERK2010FINAL.pdf>
36. Tuzel O, Porikli F, Meer P (2006) Region covariance: a fast descriptor for detection and classification
37. Ulutagay G, Nasibov E (2008) Fm-dbscan: a novel density-based clustering method with fuzzy neighborhood relations
38. Wu H, Wu W, Peng J, Zhang J (2017) A novel image retrieval algorithm based on phog and lsh. *International Journal of Computer Applications*. ARTICLES YOU MAY BE INTERESTED IN Research on image retrieval algorithm based on LBP and LSH AIP Conference A Novel Image Retrieval Algorithm based on PHOG and LSH, vol. 1864, p. 20058, 2017. <https://doi.org/10.1063/1.4992874>
39. Xiao J, Xu H, Gao H, Bian M, Li Y (2021) A weakly supervised semantic segmentation network by aggregating seed cues: the multi-object proposal generation perspective. *ACM Trans Multimed Comput Commun Appl* 17:1–19. <https://doi.org/10.1145/3419842>
40. Yan K, Wang Y, Liang D, Huang T, Tian Y (2016) Cnn vs. sift for image retrieval. *Alternative or complementary?*
41. Yang X, Gao X, Song B, Han B (2021) Hierarchical deep embedding for aurora image retrieval. *IEEE Trans Cybernet* 51(12):5773–5785. <https://doi.org/10.1109/TCYB.2019.2959261>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.