

On integrating re-ranking and rank list fusion techniques for image retrieval

K. S. Arun¹ · V. K. Govindan¹ · S. D. Madhu Kumar¹

Received: 4 July 2016 / Accepted: 3 May 2017 / Published online: 17 May 2017
© Springer International Publishing Switzerland 2017

Abstract This paper aims to unify image re-ranking and rank aggregation strategies to enhance the retrieval precision of content-based image retrieval (CBIR) systems. In general, CBIR systems are concerned with the retrieval of a set of relevant images from large repositories in response to a submitted query. The primary objective of CBIR systems is the exact ordering of database images in accordance with the presented query. To this end, we present a novel image re-ranking scheme for reordering the initial search results returned by multiple retrieval models and an efficient rank list fusion scheme to combine these refined retrieval results to achieve better performance. The re-ranking algorithm introduced in this work utilizes distance correlation coefficient to refine the search result generated by a given retrieval model. It involves two-step clustering of the initial retrieval list followed by an adaptive procedure for updating the similarity scores among images based on the created clusters. Similarly, the Particle Swarm Optimization-based similarity score fusion framework presented in this work optimally combines the retrieval results generated by multiple CBIR systems. The proposed approach is evaluated on various retrieval tasks using state-of-the-art low-level and high-level descriptors. Experimental results show that our model can significantly enhance the overall effectiveness of CBIR systems.

Keywords Image retrieval · Image re-ranking · Rank fusion

✉ K. S. Arun
iamarunks@gmail.com

V. K. Govindan
vkg@nitc.ac.in

S. D. Madhu Kumar
madhu@nitc.ac.in

¹ Department of Computer Science & Engineering, National Institute of Technology, Calicut, India

1 Introduction

Nowadays, the size of digital image repositories is growing in an exponential fashion due to the advances in data storage technologies and image capturing devices. This necessitates the design of automated models which effectively manages these large-scale image collections. Content-based image retrieval (CBIR) has emerged as a widely accepted solution to tackle this issue and it helps to organize and search digital image collections by means of their content. The notion of content either refers to visual properties (e.g., color, texture, shape etc.) or semantic information (e.g., objects present in the scene) associated with images. A ranked list of desired images is returned based on the similarity between the content description of the given query and the images already present in the database. In general, a ranking function will do the trick and the relative ordering of images in the final list indicate their degrees of relevance to the given query. However, it should be noted that certain image representation schemes and distance measures are appropriate only for some image datasets and less suitable for the rest. In other words, none of these image representation schemes and distance measures perform consistently well in all circumstances.

Recently, many post-retrieval optimization frameworks have been proposed to refine the final rankings returned by CBIR systems. These rank list optimization techniques can be grouped into three main categories: (i) approaches based on relevance feedback (RF) [1–6], (ii) fusion models [7–9] and (iii) re-ranking methods [10–15]. Relevance feedback incorporates user judgements in the retrieval process. It provides the opportunity for users to evaluate retrieval results and then automatically refine the query or similarity measure on the basis of those evaluations. Conversely, fusion models either use an aggregated feature descriptor or merge the retrieval list generated by multiple feature descriptors to

generate a consensus ranking. On the other hand, image re-ranking methods attempt to improve retrieval precision by reordering images based on the initial search results and certain auxiliary information.

RF is an online learning strategy that generates enhanced retrieval results based on the feedback from end-users regarding the relevance of images present in the originally induced ranking. The primary objective of RF is to learn the needs and preferences of end-users. To do so, the quality of the search result for a given query is judged by marking the retrieved images as being either relevant or not. Then, CBIR systems exploit this information to improve the retrieval result and a revised ordering of images is presented back to the user. This process continues until there is no further improvement in the result or the user is satisfied with the new result. In recent years, wide varieties of RF algorithms have been developed and in general these relevance feedback schemes are classified into two different classes: (i) query modification approaches and (ii) methods based on ranking function alteration. The most practiced techniques for query modification are Query Point Movement (QPM) [1] and Query EXpansion (QEX) [16]. However, global optimal results are not easily obtained with QPM- and QEX-based strategies. In contrast to query modification approaches, the second category of relevance feedback schemes modifies the ranking function by means of weighting strategies [17] or learning models [18]. In spite of their improved performance in image retrieval, feedback approaches based on ranking function alteration have many practical limitations. First, they rely on human judgements and one has to often go through several feedback iterations to achieve a better result. In practice, this is time consuming and computationally complex. Second, user has to invest extra effort in judging the relevance of images returned by the CBIR system.

The above-mentioned limitations motivated the development of unsupervised strategies in which the goodness, of multiple feature descriptors or their retrieval results, is combined during query time for better retrieval efficiency. One of the widely accepted solutions toward this direction is the fusion model, and it generally falls into two main categories namely early fusion and late fusion approaches [19]. In early fusion, multiple image descriptors are composed to form a single feature vector before indexing starts and the similarity between images is measured in terms of this aggregated feature. On the other hand, approaches based on late fusion are further split into two major groups: (i) similarity score-based rank list fusion and (ii) order based rank list fusion. In similarity score-based rank list fusion, the similarity scores of distinct image descriptors are merged by means of an aggregation function to form the final search result. The aggregation function exploits the knowledge derived from multiple rank lists for computing a more accurate ordering of images. Alternatively, order-based rank list fusion models

provide a revised retrieval result as a function of the position in which images appear in different rank lists. Since the feature characteristics and algorithmic procedures of individual methods are entirely different, feature level fusion is highly challenging. Therefore, late fusion tends to be more robust and gives better performance in terms of precision for the retrieval operation as compared to early fusion techniques.

Another widely accepted solution which enhances the retrieval effectiveness without much human intervention is image re-ranking. It is basically a post-processing analysis in which the similarity between images is recalculated with the help of an initial ranking list and some auxiliary information. In general, auxiliary information can be anything that helps the ranking function to refine the original retrieval list and is derived from the initial retrieval list in a completely unsupervised manner. This, in turn, improves the retrieval precision to a large extent. In past few years, considerable research efforts have been devoted toward the design of efficient image re-ranking algorithms. Based on how the auxiliary information is extracted from the initial ranked list, the re-ranking methods can be further classified into the following categories: clustering-based re-ranking [10,20], pseudo-relevance feedback (PRF) [21–24] and graph-based approaches [25–27]. All these approaches will be discussed in more detail later in this paper.

To summarize, lots of efforts have been made in the past to devise a variety of re-ranking and rank aggregation methods. Even though some promising results have been obtained with these formulations, there is still scope for future research to improve their retrieval precision. Moreover, most of the existing formulations disregard the possibility of combining the advantages of these two methods for the retrieval task. To this end, novel approaches for image re-ranking and rank aggregation are proposed. Further, the feasibility of combining re-ranking and rank aggregation methods to obtain more accurate retrieval results is also explored. The formulated model works in the following fashion. First, the newly introduced image re-ranking algorithm is applied separately to the retrieval list returned by state-of-the-art image descriptors and then these refined lists are combined by the proposed rank aggregation algorithm. The main contributions of this paper are:

1. A distance correlation coefficient-based image re-ranking scheme to update the retrieval list generated by a given CBIR system.
2. A Particle Swarm Optimization-based rank aggregation framework to aggregate the retrieval list generated by multiple CBIR systems.
3. An approach for combining the results of re-ranking and rank aggregation aiming at improving the effectiveness of CBIR systems.

The remainder of this paper is organized as follows. Sect. 2 summarizes the related works in re-ranking and rank aggregation-based image retrieval. Section 3 will introduce the notations and definitions used in this paper. The proposed image re-ranking scheme is described in Sect. 4. The formulation of the PSO-based rank list fusion framework is explained in Sect. 5. The combination strategy for integrating the results of the proposed re-ranking and rank aggregation algorithms is discussed in Sect. 6. The experimental evaluation of the proposed image re-ranking and rank list fusion algorithms are presented in Sect. 7. Section 8 enumerates the research outcome and also outlines the directions for future research.

2 Prior work

This section summarizes the state-of-the-art research in image retrieval using re-ranking and rank aggregation-based strategies. In Sect. 2.1, the existing approaches for image re-ranking are discussed in detail and Sect. 2.2 outlines various rank list fusion methods.

2.1 Image re-ranking

A comprehensive review of various image re-ranking techniques such as pseudo-relevance feedback, graph-based and clustering-based approaches are provided in this section. Pseudo-relevance feedback based re-ranking is entirely based on the assumption that only the top-ranked images in the retrieval list are considered as relevant to the given query. These top-ranked images are termed as pseudo-relevant. This is in contrast to RF-based rank list refinement where users explicitly provide feedback by labeling the results as relevant or irrelevant. Then these pseudo-relevant images can be either used to train a statistical model by which images in the original retrieval list can be re-arranged according to the confidence scores yielded by the learned model or provided as feedback to the retrieval system for query reformulation. It should be noted that the pseudo-relevance assumption still preserves the unsupervised nature of the re-ranking process. To this end, Shen et al. [14] proposed k-NN re-ranking which automatically refines the initial rank list using the k-nearest neighbors of the given query. Alternatively, Qin et al. [28] take advantage of k-reciprocal nearest neighbors to identify the set of relevant images for re-ranking. However, the main limitation of this approach is how to select the pseudo-relevant images from the initial ranked list and how to efficiently employ these images for the re-ranking task.

More recently, graph-based approaches for image re-ranking are gaining increasing popularity. In graph-based re-ranking, a similarity graph $G = (V, E)$ is constructed

over the initial retrieval list with each node $v \in V$ corresponds to an image in the data set and an edge $e \in E$ denotes the similarity among images. The graph G is created in such a way that visually similar images are neighbors in G and their similarity scores are close to each other. Then, the technique of link analysis can be employed to find out the contextual patterns embedded in G to re-order the original retrieval list. Jing et al. [29] applied PageRank algorithm on image similarity graph to re-arrange the initial retrieval list. They use the stationary probability of the random walk as an improved similarity score for the re-ranking operation. In a similar fashion, Hsu et al. [11] proposed the notion of context graph and employed random walk along this newly formulated context graph to re-rank the initial search result of large-scale product image dataset. On the contrary, Tian et al. [13] formulated re-ranking as a global optimization problem within a Bayesian framework. That is, they modeled the re-ranking problem from the probabilistic perspective and derived an optimal re-ranking function based on Bayesian analysis.

On the other hand, clustering-based image re-ranking algorithms relies on the fact that an initial retrieval list can be further partitioned into relevant and irrelevant ones using appropriate clustering algorithms. After this preliminary grouping, images in the cluster that are similar to the given query are placed on top of the retrieval list to enhance the retrieval precision. In this direction, Park et al. [10] employed Hierarchical Agglomerative Clustering (HAC) to analyze the initial retrieval list and the ordering of images in individual group are adjusted in accordance with the distance of the query image to the resulting clusters. However, the clustering-based approaches have the following limitation: (i) how to perform clustering in the initial ranked list and (ii) how to rank the clusters and images within each cluster.

To overcome these limitations, several advanced strategies have been proposed and the most notable among them is image correlation-based re-ranking techniques. The traditional re-ranking methods only consider pairwise image similarity, and they completely ignore the correlation among images in the whole dataset. However, correlation-based approaches aims to improve the retrieval effectiveness by replacing the pairwise image similarity calculation using global affinity measures which incorporate the correspondence among all the images in the database. In this regard, graph transduction [30], diffusion process [31], affinity learning [32] and context-based algorithms [15, 33, 34] have been introduced. Among all these approaches, context-based re-ranking is more prominent and it requires special attention.

While judging image similarity, context-based re-ranking algorithms integrate various sources of supplementary information. Initially, Pedronette and Torres [33] proposed Distance Optimization Algorithm (DOA) for image re-ranking. It is basically an iterative clustering approach based on the distance correlation measure. In essence, DOA exploits the

fact that if two images are similar their distances to the rest of the images in the dataset and the corresponding retrieval lists when these two images supplied as query should be identical. Later on, RL-Sim algorithm is introduced by Pedronette and Torres [15] for image re-ranking. It is basically an iterative approach where the distance between images is updated in each step based on the similarity of the retrieval lists of the database images. More recently, Pedronette et al. [34] developed Reciprocal K-NN Graphs Based Manifold Learning (RKNN-ML) algorithm for image re-ranking. In their approach, the affinity between ranked lists of database images is encoded in the form of k-reciprocal neighborhood graph and manifold learning is further used to update the similarity between images.

The re-ranking algorithm proposed in this paper incorporates contextual information for reordering an initial retrieval list. The contextual information is encoded in the form of distance correlation coefficient. The correspondence between two images is determined on the basis of their similarity scores to the rest of the images in the database. Distance correlation coefficient is a numerical measure to characterize the strength of the correspondence of similarity scores between images. Therefore, distance correlation coefficient-based image re-ranking scheme updates the similarity score between images in an adaptive fashion by considering the correlation statistics. The proposed algorithm has another important advantage that it performs equally well in low-level and high-level descriptor-based image retrieval systems.

2.2 Rank list fusion

The objective of rank list fusion is the aggregation of outputs from different but complementary retrieval models to generate a more comprehensive retrieval result. In conventional CBIR systems, the search result is generated based on the similarity score computed from a single feature descriptor. On the contrary, in rank list fusion, an integrated ordering of the search results from multiple retrieval models is accomplished by means of a fusion algorithm. Generally, the fusion algorithm is designed in such a way that optimizes the overall retrieval performance. The fusion algorithms mainly use the following information to get a consensus ranking: (i) the rank positions assigned to images in individual retrieval list or (ii) the similarity scores of the database images returned by different models.

Rank position-based fusion makes use of order information of images from various retrieval lists to realize rank list aggregation. Early efforts in order-based rank list fusion completely depend on heuristic algorithms. For example, the Borda Count (BC) method [35] in which images with the highest rank on each retrieval list gets n votes, where n is the size of the image collection. Each subsequent rank position gets one vote less than the previous. The votes across

multiple rank lists are then summed up, sorted and presented as the final aggregated list. In contrast to this, the Reciprocal Rank Fusion (RRF) [9] scheme employs the mean of the harmonic means of the ranking information across different models to generate the final result. Nowadays, probabilistic models on permutations, such as the Mallows model [36] and the Plackett–Luce model [37], have been widely used to solve the problem of order-based rank list fusion. Most of these approaches rely on a probability distribution built over the space of different rankings of images to get an enhanced retrieval result. Another good practice is the use of Kemeny Optimal Aggregation (KOA) [7] which tries to optimize the average Kendall–Tau distance between the fused result and the original retrieval lists. Kendall–Tau distance counts the pairwise disagreements between two retrieval lists. In practice, position-based rank aggregation methods are computationally efficient, but retrieval precision of a desired level is not achieved by any of these models. Therefore, we focus on similarity score-based aggregation in this paper.

Similarity score-based fusion follows a different strategy to combine the search results returned by different retrieval models. An earliest attempt toward direction is the Markov chain-based approaches. Here, images belongs to various rank lists are represented by the nodes of a directed graph and the transition probabilities among these nodes are defined in terms of the relative ranking of images in various retrieval lists. Then, a fused ranking is obtained by computing a stationary distribution on the Markov chain. Dwork et al. [7] proposed several Markov chain-based methods for rank aggregation, namely, MC1, MC2, MC3, and MC4. These methods differ from each other in the way the transition probabilities are calculated. It should be noted that all these models perform reasonably well for rank lists of varying size.

Later on, graph fusion techniques [26,27] have been widely adopted for similarity score-based rank list fusion. In graph-based fusion, the search results from individual retrieval models are formally represented with a graphical structure known as image graph. In general, an image graph is a weighted undirected graph where each node represents an image and the edges encode the similarity or affinity between images. For each retrieval model, an image graph centered on the query is constructed whose remaining vertices correspond to the images in the retrieval list. Edges are included in the graph based on the pairwise affinity between images. Finally, these multiple image graphs are merged to form a single one and an efficient ordering of the candidate images is obtained with the use of graph-based ranking algorithms. Wang et al. [26] formulated the task of rank fusion as an optimization problem involving normalized graph Laplacian regularization term. An iterative optimization procedure known as manifold ranking is then used to estimate relevance score of all the images in the dataset. Zhang et al. [27] calculated the edge weights of image graphs by means of

Jaccard similarity coefficient and multiple image graphs are then fused by simply accumulating the edge weights and link analysis is performed to get the final ordering of the candidate images.

More recently, numerous unweighed and weighted rank list fusion models have been introduced. For instance, Fox and Shaw [38] introduced a family of unweighed combination strategies for rank list fusion such as CombSUM, CombMIN, CombMAX, CombANZ and CombMNZ. CombSUM arrange images based on the sum of the similarity scores of the individual models, while CombMIN and CombMAX strategies consider the maximum and minimum scores secured by individual images for preparing the final ranking. Conversely, the fused similarity in CombMNZ is derived as the sum of the scores generated by individual model (i.e., CombSUM) multiplied by the total number of retrieval models. CombANZ is similar to CombMNZ except that, instead of multiplying, we divide CombSUM by the total number of retrieval models.

In contrast to the above-mentioned approaches, Jain and Vailaya [39] introduced a weighted combination of the shape- and color-based image descriptors for the construction of an improved rank list. Later on, a detailed analysis of various similarity score-based rank list fusion schemes is reported by Depeursinge and Muller [40]. They established the fact that if reasonable weights, for similarity scores arrived at with different retrieval frameworks, have been obtained, then weighted model is the best method for all situations. However, most of the existing rank list fusion algorithms give equal weights to the similarity scores returned by the constituent models. This assumption does not always hold true. In practice, for a given query, the retrieval list generated by a particular model is sometimes found to be superior than the rest. In other words, the significance of each retrieval list is query specific. Hence, it is not reasonable to equally rate the ranking lists generated by multiple retrieval models. Weighted adaptive fusion models, where the similarity scores returned by different retrieval models are assigned different weights based on the given query can somehow overcome this issue. In this paper, we analyze the retrieval results generated by different models in response to the submitted query to infer reasonable fusion weights for the corresponding similarity scores. To find a preferred solution, an optimization problem is formulated and is resolved using a PSO-based algorithm.

3 Notations and definitions

The basic notations used throughout this paper and the formal definitions of image re-ranking and rank aggregation problems are provided in this section.

Let $\mathbb{C} = \{I_1, I_2, \dots, I_n\}$ be the collection of images in the given dataset and $f \in \mathbb{R}^d$ be the feature descriptor used to

characterize individual images in \mathbb{C} . Let $\mathbb{S} : f \times f \rightarrow \mathbb{R}$ denote the similarity function used to measure the correspondence between images in \mathbb{C} . Then, the similarity scores $S(I_j, I_k)$ among all pairs of images $(I_j, I_k) \in \mathbb{C}$ yield a similarity matrix $S_{n \times n}$ and is finally used to generate a ranked list of images R_q in response to a given query I_q . The retrieval list R_q can be viewed as a permutation of the images in the dataset \mathbb{C} where an image I_j is placed on top of another image I_k if and only if $S(I_q, I_j) < S(I_q, I_k)$.

With this basic introduction, an image re-ranking procedure can be formally defined as a function $\Psi(\cdot)$ that accepts as input the initial similarity matrix S_{init} of the retrieval model under consideration and provides a more reasonable similarity matrix S_{new} as follows:

$$S_{\text{new}} = \Psi(S_{\text{init}}) \quad (1)$$

A refined retrieval list is then obtained with this new similarity matrix S_{new} . In practice, the re-ranking function $\Psi(\cdot)$ exploits certain auxiliary information along with the initial similarity matrix S_{init} to infer a more effective similarity matrix S_{new} .

Let $F = \{f_1, f_2, \dots, f_m\}$ be the set of m image descriptors for the given image collection \mathbb{C} and $\Omega = \{S_1, S_2, \dots, S_m\}$ be the corresponding similarity matrices, then for a given query I_q , a rank aggregation function $\Phi(\cdot)$ unifies these similarity matrices to form an aggregated similarity matrix S_{agg} as stated below:

$$S_{\text{agg}} = \Phi(\Omega) \quad (2)$$

The aggregation function $\Phi(\cdot)$ is defined as $\Phi : S_1 \times S_2 \times \dots \times S_m \rightarrow S_{\text{agg}}$ where S_{agg} is the unified similarity matrix of the dataset \mathbb{C} for the given query I_q . Finally, the retrieval system returns a better search result on the basis of the aggregated similarity matrix S_{agg} .

4 Distance correlation coefficient-based image re-ranking

The proposed image re-ranking scheme relies on the fact that retrieval effectiveness of CBIR systems can be considerably enhanced by exploiting the contextual information hidden in the similarity matrix. In general, to compute the similarity scores among images, only pairwise analysis is performed and in most cases the relationship among all the images in the database is completely ignored. The distance optimization algorithm (DOA) proposed by Pedronette and Torres [33] update the distance between images based on the correlation of the similarity scores of their nearest neighbors. In practice, DOA updates the similarity scores among images based on the correspondence of their ranked lists. For this, an iterative

Algorithm 1 Distance correlation coefficient-based image re-ranking scheme

Input: The query image I_q , the initial retrieval list R_q and the corresponding similarity matrix S_{init} , cluster size K

```

1: Initialize  $P_1 = [I_q]$ ;  $P_2 = [\phi]$ ;  $Cl_1 = [\phi]$ ;  $Cl_2 = [\phi]$ ;  $Cl_3 = [\phi]$ ;
2: for all  $I_d \in \mathbb{C}$  do
3:   if  $dCor(I_q, I_d) \geq \theta$  then
4:      $P_1 = [P_1 \mid I_d]$ 
5:   else
6:      $P_2 = [P_2 \mid I_d]$ 
7:   end if
8: end for
9: for all  $I_r \in R_q$  do
10:  if  $I_r \in P_1$  then
11:     $Cl_1 = [Cl_1 \mid I_r]$ 
12:  else if ( $I_r \in P_2$ ) AND ( $r < K$ ) then
13:     $Cl_2 = [Cl_2 \mid I_r]$ 
14:  else
15:     $Cl_3 = [Cl_3 \mid I_r]$ 
16:  end if
17: end for
18: for all  $(I_j, I_k) \in Cl_1$  do
19:   $S_{new}(I_j, I_k) = dCor(I_j, I_k) * S_{init}(I_j, I_k)$ 
20: end for
21: for all  $(I_j, I_k) \in Cl_2$  do
22:   $S_{new}(I_j, I_k) = \left(1 + \frac{1}{1-dCor(I_j, I_k)}\right) * S_{init}(I_j, I_k)$ 
23: end for
24: for all  $(I_j, I_k) \in Cl_3$  do
25:   $S_{new}(I_j, I_k) = \frac{1}{dCor(I_j, I_k)} * S_{init}(I_j, I_k)$ 
26: end for
27: return  $S_{new}$ 

```

clustering approach is employed where the correspondence of retrieval lists is measured in terms of Pearson's correlation coefficient.

The proposed image re-ranking scheme is inspired from DOA with the following modifications. Primarily, the correspondence of the similarity score distribution of any two images is measured in terms of the distance correlation coefficient. Secondly and most importantly, it requires one pass clustering rather than multiple iterations to update the similarity scores. The update operation is performed in an adaptive manner using the distance correlation coefficient.

The proposed image re-ranking scheme involves two clustering steps and a query-dependent update procedure for modifying the similarity score as depicted in Algorithm 1. The first clustering step is single pass in nature, and it splits the database images into two disjoint partitions (steps 1, 2 of Algorithm 1). For this, distance correlation-based statistics is employed. The set of images that are more similar to the given query are placed into the first partition and the rest of the images are placed into the second partition. The second clustering step is also a one-pass approach, and it groups the images in the retrieval list into three clusters based on the partitions created in the previous step (steps 1, 2 of Algorithm 1). Then, update rules are defined for improving the similarity score and hence to perform the re-ranking opera-

tion (steps 1, 2 of Algorithm 1). In summary, the proposed image re-ranking scheme involves the following steps:

1. Partitioning the database images into two disjoint subsets.
2. Adaptively updating the similarity score.

Before explaining these two steps in detail, we introduce the notion of distance correlation coefficient in the coming section.

4.1 Distance correlation coefficient

The proposed distance correlation coefficient-based image re-ranking scheme is fully dependent on the distribution of similarity scores of images. Consider two reference images I_j, I_k and a two-dimensional plane where the X -axis denotes the similarity score of database images with respect to I_j and the Y -axis denotes the similarity score of database images with respect to I_k . Then, the position of a database image $I_d \in \mathbb{C}$ on the $X - Y$ plane is defined by an ordered pair $(S(I_j, I_d), S(I_k, I_d))$ where $S(I_j, I_d)$ and $S(I_k, I_d)$ represents the similarity of I_d with regard to the reference images I_j and I_k .

Figure 1 depicts the similarity score distribution of INRIA Holiday dataset [41] with respect to two randomly selected images that are close to each other. The similarity between images is estimated using Steerable Pyramid based Texture Feature (SPTF) [42]. From this example, it is well understood that the similarity score distribution of two similar images is linear in nature. In other words, if the reference images are identical, then they have equal distances to the rest of the database images. In a similar fashion, Fig. 2 depicts the similarity score distribution of two reference images that are not identical. The similarity between images is again estimated on the basis of SPTF [42], and it can be inferred that the similarity score distribution of the reference images is nonlinear when they are dissimilar.

This paper aims to incorporate the above-mentioned image similarity distribution in the re-ranking process. In order to mathematically characterize the functional relationship of the similarity score vectors of two images, we implement a relatively new statistical measure called distance correlation coefficient (dCor) [43]. The name distance correlation comes from the fact that it uses the distances between observations as part of its calculation. Let Z^j and Z^k denotes the similarity matrices S_{I_j} and S_{I_k} of the two images I_j and I_k in vector form, then the distance correlation coefficient (dCor) is formally defined as:

$$dCor(Z^j, Z^k) = \sqrt{\frac{dCov^2(Z^j, Z^k)}{dCov^2(Z^j, Z^j) dCov^2(Z^k, Z^k)}} \quad (3)$$

where $dCov^2(Z^j, Z^k)$ is the distance covariance and is expressed as:

Fig. 1 Similarity score distribution of INRIA Holiday data set [41] with respect to two similar images

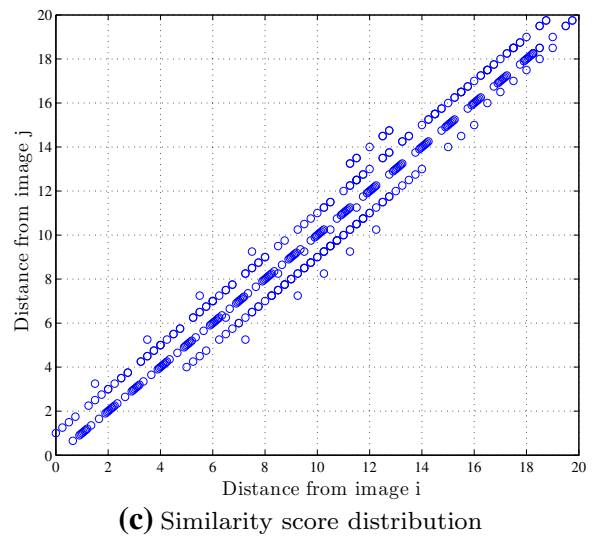
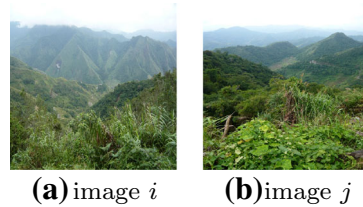
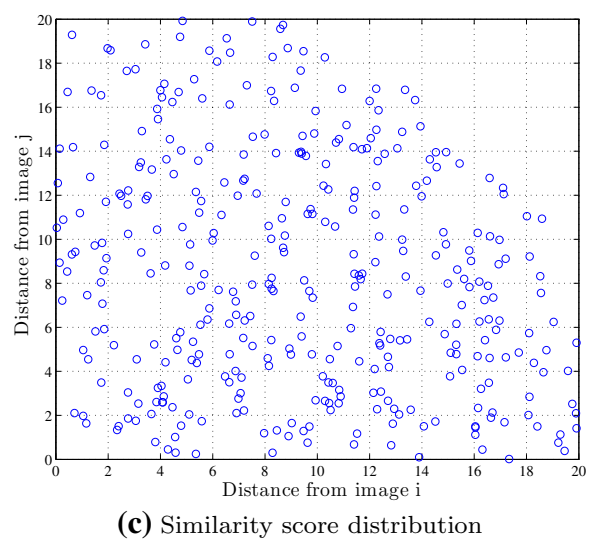
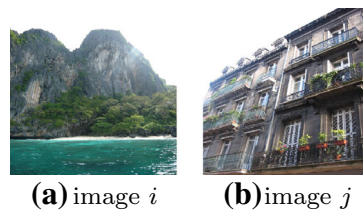


Fig. 2 Similarity score distribution of INRIA Holiday data set [41] with respect to two dissimilar images



$$dCov^2(Z^j, Z^k) = \frac{1}{n^2} \sum_{p=1}^n \sum_{q=1}^n Z_{pq}^j Z_{pq}^k \tag{4}$$

Distance correlation coefficient always takes value in the range [0,1] and $dCor = 0$ only if Z^j, Z^k are independent and $dCor = 1$ when Z^j, Z^k are identical. With this basic introduction to distance correlation coefficient, the rest of this section explains the major steps involved in the proposed image re-ranking scheme.

4.2 Partitioning the database images into two disjoint subsets

Let I_q be the submitted query and \mathbb{C} be the given image collection then, the first step of the proposed image re-ranking algorithm partitions \mathbb{C} into two disjoint subsets in accordance with the distance correlation coefficient calculated between the query and all the images in the collection. Once the distance correlation coefficient $dCorr(I_q, I_d)$ between the query

I_q and an image $I_d \in \mathbb{C}$ is calculated, a membership assignment function of the following form is defined

$$\eta(I_q, I_d) = \begin{cases} 1, & \text{if } dCorr(I_q, I_d) \geq \theta \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where θ denotes a threshold value and is to be tuned according to the dataset under evaluation. Based on the value of the membership assignment function, the initial retrieval list is partitioned into two disjoint subsets P_1 and P_2 as follows:

1. if $\eta(I_q, I_d) == 1$, then assign I_d to the first partition P_1
2. if $\eta(I_q, I_d) == 0$, then assign I_d to the second partition P_2 .

Thus, the images assigned to partition P_1 are highly correlated with the given query I_q and those placed in P_2 are likely to be less correlated with I_q .

4.3 Adaptively updating the similarity score

In this step, the information obtained from the previously created image partitions together with the correlation statistics are efficiently utilized to update the initial similarity scores among images. Let R_q be the retrieval list generated by the given query I_q and $I_r \in R_q$, then the update operation is performed by assigning the images present in R_q to three clusters as follows:

1. An image $I_r \in R_q$ is placed in cluster Cl_1 if I_r belongs to the partition P_1 .
2. An image $I_r \in R_q$ is placed in cluster Cl_2 if $I_r \in P_2$ and the index of I_r in the original retrieval list R_q is such that $r < K$, where K is a user defined constant and it determines the size of the cluster Cl_2 .
3. An image $I_r \in R_q$ is placed in cluster Cl_3 if $I_r \notin Cl_1$ and $I_r \notin Cl_2$.

Next, three different update rules are defined to improve the similarity scores of images belonging to clusters Cl_1 , Cl_2 and Cl_3 . Let $S_{init}(I_j, I_k)$ be the initial similarity score between the image pairs (I_j, I_k) , then for each cluster an updated similarity score $S_{new}(I_j, I_k)$ is calculated by the following rules:

1. If $(I_j, I_k) \in Cl_1$, then $S_{new}(I_j, I_k) = dCor(I_j, I_k) * S_{init}(I_j, I_k)$
2. If $(I_j, I_k) \in Cl_2$, then $S_{new}(I_j, I_k) = \left(1 + \frac{1}{1-dCor(I_j, I_k)}\right) * S_{init}(I_j, I_k)$
3. If $(I_j, I_k) \in Cl_3$, then $S_{new}(I_j, I_k) = \frac{1}{dCor(I_j, I_k)} * S_{init}(I_j, I_k)$

Thus, the initial similarity scores are updated in an adaptive manner based on the distance correlation coefficient. When the image pairs under examination (I_j, I_k) belongs to cluster Cl_1 , then they are closer to the given query and the similarity score modification involves decreasing their pairwise distances. Therefore, an updated version of the similarity score is obtained by multiplying the original score with a value equal to $dCor(I_j, I_k)$. As already mentioned the values of $dCor(I_j, I_k)$ ranges in $(0, 1]$, the distances among similar images are considerably declined. On the contrary, when the image pair belongs to Cl_2 , then their similarity to the given query is uncertain and the value of the multiplication constant is then adaptively determined based on the distance correlation coefficient. In this case, the value of the multiplication constant is set to $1 + \frac{1}{(1-dCor(I_j, I_k))}$ and it always yields a weight greater than 1. All the remaining image pairs are considered to be dissimilar to the given query, and in this case the original similarity scores are multiplied

by a factor $\frac{1}{dCor(I_j, I_k)}$. At the end, images in the dataset are re-ordered based on the updated similarity scores.

5 The proposed rank list fusion scheme

As mentioned earlier, each image representation scheme along with its distance measure might be complementary in nature and will have its own merits and demerits. Therefore, fusing the retrieval lists generated by these independent and heterogeneous models is expected to yield a better result than each of the strategies in isolation. This motivates us to develop new methods for fusing the search results generated by multiple retrieval models for a better retrieval precision.

Even though several approaches to enhance image retrieval performance based on similarity score fusion have been reported, most of them utilize a query independent fusion scheme. That is, a learned model is directly applied to the retrieval lists generated by multiple models by ignoring the fact that only a few feature descriptors will have a significant impact on the final retrieval result. In other words, the similarity scores returned by all the feature descriptors are not equally important and its significance varies according to the given query. To overcome this limitation, this paper proposes a novel rank list fusion scheme which learns optimal fusion weights for the similarity scores generated by different feature descriptors based on the given query. With the proposed scheme, the aggregated similarity score $S_{agg}(I_q, I_d)$ corresponding to the given query I_q and a database image I_d is calculated as:

$$S_{agg}(I_q, I_d) = \frac{\sum_{i=1}^n w_{f_i} S_{f_i}(I_q, I_d)}{\sum_{i=1}^n w_{f_i}} \quad (6)$$

where $S_{f_i}(I_q, I_d)$ is the similarity score returned by the i -th feature vector f_i and w_{f_i} is the weight corresponding to the similarity score $S_{f_i}(I_q, I_d)$. Since the effectiveness of rank list fusion fully depends on the choice of n -dimensional fusion weight vector $W = [w_{f_1}, w_{f_2}, \dots, w_{f_n}]$, a better retrieval list is realized by finding optimal values to fusion weights w_{f_i} . An optimization problem is formulated to infer the optimal fusion weights in accordance with the submitted query and the next section elaborates how the task of finding query adaptive fusion weights can be formulated as an optimization problem.

5.1 Problem definition

For a given query, the major concern while performing rank list fusion is the assignment of reasonable weights to the similarity scores returned by various feature descrip-

tors. This section provides the formal definition of the objective function to be optimized by the rank list fusion scheme to infer query-dependent fusion weights w_{f_i} . Let $L = \{L^1, L^2, \dots, L^t\}$ be the set of aggregated retrieval lists corresponding to t different values of the fusion weight vector $\{W^1, W^2, \dots, W^t\}$. For the sake of simplicity, let us consider the top K images from each of the n fused retrieval list for evaluation. This results a total of $t \times K$ retrieved images corresponding to t different fusion weight vectors. Then, the quality of each of these fused retrieval list is judged in terms of the membership degree of its top k images in the rest of the $(t - 1)$ retrieval lists. This can be mathematically formulated as follows.

Let δ_{pj}^i be an indicator function which denote whether the p -th image (where $p \leq K$) I_p^i of the i -th aggregated retrieval list L^i is also present within the top- K position of the j -th aggregated list L^j . That is,

$$\delta_{pj}^i = \begin{cases} 1, & \text{if } I_p^i \in L_j \\ 0, & \text{if } I_p^i \notin L_j \end{cases} \quad (7)$$

Then, the sum of the membership degree of the p -th image (where $p \leq K$) I_p^i of L^i across all the retrieval list L is given by:

$$M_p^i = \sum_{j=1}^t \delta_{pj}^i \quad (8)$$

For all the images within the top K position of the i -th retrieval list L^i , the overall membership degree is then defined as:

$$M^i = \sum_{q=1}^K M_q^i \quad (9)$$

Finally, the normalized version of the overall membership degree of the aggregated retrieval list L^i is calculated as:

$$H^i = \frac{M^i}{\sum_{r=1}^t M^r} \quad (10)$$

Therefore, the optimal fusion weights with respect to the given query I_q is the one that maximizes the normalized membership degree of the set of all aggregated retrieval lists considered for evaluation and is mathematically defined as:

$$\text{maximize } H^i = \frac{M^i}{\sum_{r=1}^t M^r}, \quad \forall i \in 1, 2, \dots, t \quad (11)$$

As the value of H^i is high, the images in the fused retrieval list L^i occupies a greater proportion of all the $t \times K$ images

considered for evaluation. Hence, the retrieval list L^i is considered as the most prominent as compared to the rest of the $(t - 1)$ lists in the set L and the corresponding weight vector W^i is regarded as the optimal fusion weights for the given query.

5.2 PSO-based rank list fusion algorithm

In this paper, the task of finding weight values corresponding to the similarity scores returned by different image descriptors is formulated as a numerical optimization problem. Over the years, various approaches have been proposed to solve a wide range of numerical optimization problems and it is required most of the classical optimization techniques to comply with the structure of the objective function intended to be solved. In practice, if the derivative of the objective function with respect to the variable to be optimized cannot be calculated as in the case of Eq. (11), then it gets difficult to find an optimal solution by means of the classical approaches. In such situations, it is a common practice to use metaheuristic algorithms. The most widely used metaheuristic algorithms in scientific applications are: Genetic Algorithm (GA)[44], Particle Swarm Optimization algorithm (PSO)[45], Differential Evolution (DE) [46], Artificial Bee Colony (ABC) algorithm [47] and Cuckoo Search Algorithm (CSA) [48].

More recently, Wahab et al. [49] provided a comprehensive evaluation of the performance of various meta heuristic algorithms in solving a set of thirty benchmark functions. In their experiments, the benchmark selected for evaluation differs in their characteristics and it includes unimodal, multimodal, separable and inseparable functions. The evaluation results thus obtained clearly indicated the superiority of PSO in solving optimization problems involving unimodal functions. It has been observed that PSO outperformed or performed equally to the best algorithm in eleven out of the twelve unimodal functions selected as benchmark. These results prompted us to choose PSO for inferring optimal fusion weights that efficiently combine the similarity scores returned by multiple image descriptors. To keep things simple, a brief overview of Particle Swarm Optimization (PSO) is provided in the next section and then the proposed rank list fusion framework based on PSO is discussed.

5.2.1 Overview of PSO

Particle Swarm Optimization (PSO) is a population-based stochastic optimization technique developed by Kennedy et al. [45] and is motivated by the social behavior perceived in flocks of birds and schools of fish. In bird flocking or fish schooling, there exist a leader to direct the group forward and all the other members of the group will follow the leader. In other words, individuals in the group exchange previous

experience and accordingly adjust their position so that they can move toward the objective. The same concept is adopted by PSO while searching an optimal solution for a given optimization problem.

In PSO, there exist a population (or *swarm*) of potential solutions (or *particles*) to the problem under consideration and in successive iterations each particle in the swarm moves in a multi-dimensional solution space in search for a global optimum. The movement of the swarm in the solution space is mainly governed by two factors namely the past experience of individual particles and the knowledge gained from the current best particle of the entire swarm. All the particles inside the swarm are evaluated based on their fitness. A fitness function of the form $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined for this purpose. In fact, the fitness function accepts a particle as input in the form of a vector of real numbers and yields a real number as output which specifies the fitness of the particle considered for evaluation. The basic steps involved in PSO can be summarized as follows:

1. *Swarm initialization* as the first step, an initial swarm of particles is created in the solution space. In general, the nature of the optimization problem decides the number of particles in the swarm. Each particle i will have a position $p_i \in \mathbb{R}^n$ and a velocity $v_i \in \mathbb{R}^n$ in the search space. In general, the position p_i of the i -th particle is initialized with a uniformly distributed random vector, i.e., $p_i \sim U(s_{lo}, s_{up})$, where s_{lo} and s_{up} are the lower and upper boundaries of the search space. Similarly, the velocity of the i -th particle is initialized as $v_i \sim U(-|s_{up} - s_{lo}|, |s_{up} - s_{lo}|)$.
2. *Iteratively update the swarm* in every iteration, each particle is updated based on its best-known position in the search space as well as the entire swarm's best-known position. The former is known as previous best position ($pBest$) and the later is termed as global best position ($gBest$). That is, each particle communicates with its neighbors about its position, memorizes its best position so far and also knows the position of the highest performing neighbor. Once $pBest$ and $gBest$ are obtained, a particle updates its velocity and position as follows:

$$v_i^{(t+1)} = v_i^{(t)} + C_1 * R_1 * (pBest - p_i^{(t)}) + C_2 * R_2 * (gBest - p_i^{(t)}) \quad (12)$$

$$p_i^{(t+1)} = p_i^{(t)} + v_i^{(t+1)} \quad (13)$$

where C_1 is the cognition parameter and C_2 is the social parameter which serves as acceleration coefficients that are conventionally set to a fixed value between 0 and 2. R_1 and R_2 are random numbers within the range (0, 1). $v_i^{(t)}$ and $v_i^{(t+1)}$ represents the velocity of the particle i

at iteration t and $t + 1$. Similarly, $p_i^{(t)}$ and $p_i^{(t+1)}$ corresponds to the position of the particle at iteration t and $t + 1$. Besides this, the fitness value of all particles are calculated in each iteration and the values of $pBest$ and $gBest$ are then updated if particles with better position or global best position is obtained.

3. *Termination* steps (2) is repeated iteratively until an adequate fitness is reached or a maximum number of iterations is performed. A predefined error value is initially provided to check whether an adequate fitness is attained or not. To do so, the difference in fitness function values of successive iterations is calculated and if it is found to be less than or equal to the given error, then the entire procedure is terminated with the value of $gBest$ as the optimal solution.

The pseudo-code for the above procedure is depicted in Algorithm 2.

Algorithm 2 Particle swarm optimization

Input: Upper and lower bound of the solution space. N - the number of particles in the swarm. C_1, C_2

- 1: Initialize the position and velocity of each particle in the swarm.
 - 2: Initialize $pBest$ and $gBest$ values.
 - 3: **repeat**
 - 4: **for each particle do**
 - 5: Calculate the fitness value of the *particle*
 - 6: **if** the fitness value of the *particle* is better than the fitness of $pBest$ **then**
 - 7: set the current position of the *particle* as its new $pBest$
 - 8: **end if**
 - 9: **end for**
 - 10: $gBest \leftarrow$ among all the particles in the swarm, select the one with highest fitness value
 - 11: **for each particle do**
 - 12: Update the velocity of the *particle* according to Eq. 12
 - 13: Update the position of the *particle* according to Eq. 13
 - 14: **end for**
 - 15: **until** minimum error criteria is satisfied or maximum iterations is reached
-

5.2.2 Swarm initialization

To perform PSO-based rank list fusion, we initially need to draw a set of N particles from an n -dimensional search space of fusion weights. In PSO-based optimization, the final solution greatly depends on the number of particles, their initial position and velocity. In this paper, the number of particles in the population N is set to a reasonably large value with the aim of deriving optimal fusion weights quickly. To initialize the position of individual particles in the swarm, the solution space is originally divided into N equal regions. Then, the centroids of each such region are taken as the starting position of individual particles. The velocity of each particle is ini-

tialized as a uniformly distributed random vector in the range $[-|s_{up} - s_{lo}|, |s_{up} - s_{lo}|]$, where s_{lo} and s_{up} are the lower and upper bounds of the solution space. The best-known positions ($pBest$) of individual particles are initialized with the values of their starting position. From these values of $pBest$, the best candidate is chosen and assigned as the value of global best position ($gBest$).

5.2.3 Optimal weight finding

Algorithm 3 depicts the proposed PSO-based rank list fusion scheme for finding optimal fusion weights. It is basically an iterative procedure and works by simultaneously preserving many particles in the search space. At first, the velocity and position of individual particles in the swarm as well as $pBest$ and $gBest$ are initialized as per the procedure described in

Algorithm 3 PSO-based rank list fusion scheme

Input: (s_{lo}, s_{up}) - lower and upper boundaries of the solution space. N -the number of particles in the swarm, acceleration coefficients C_1 and C_2 , the fitness function $f(\cdot)$ to be optimized, difference in cost function values in successive iteration ϵ , the maximum number of iterations T .

```

1: Initialize  $t=1$ ;
2: Divide the solution space ( $s_{lo}, s_{up}$ ) into  $N$  equal regions and initialize the position of each particle with the centroid of these regions.
3: Initialize the global best position of the entire swarm as  $gBest \leftarrow U(s_{lo}, s_{up})$ .
4: for each particle  $i \in \{1, 2, \dots, N\}$  do
5:   Initialize the velocity  $v_i^{(t)}$  of the particle  $i$ 
6:    $v_i^{(t)} \sim U(-|s_{up} - s_{lo}|, |s_{up} - s_{lo}|)$ 
7:   Initialize the best-known position  $pBest$  of particle  $i$ 
8:    $pBest_i \leftarrow p_i$ 
9:   if ( $f(pBest_i) > f(gBest)$ ) then
10:     $gBest \leftarrow pBest_i$ 
11:   end if
12: end for
13: repeat
14:   for each particle  $i \in \{1, 2, \dots, N\}$  do
15:    select random numbers  $R_1 \sim U(0, 1)$  and  $R_2 \sim U(0, 1)$ 
16:     $v_i^{(t+1)} = v_i^{(t)} + C_1 * R_1 * (pBest - p_i^{(t)}) + C_2 * R_2 * (gBest - p_i^{(t)})$ 
17:     $p_i^{(t+1)} = p_i^{(t)} + v_i^{(t+1)}$ 
18:    if  $f(p_i^{(t+1)}) > f(pBest_i)$  then
19:      Update the particle's best-known position
20:       $pBest_i \leftarrow p_i^{(t+1)}$ 
21:    end if
22:    if  $f(pBest_i) > f(gBest)$  then
23:      Update global best position of the entire swarm:  $gBest \leftarrow pBest_i$ 
24:    end if
25:   end for
26:   if ( $f^{(t)}(gBest) - f^{(t-1)}(gBest) \leq \epsilon$ ) then
27:     break;
28:   end if
29:    $t = t + 1$ ;
30: until  $t = T$ 
31: return  $gBest$ 

```

Sect. 5.2.2 (step 2–7 of Algorithm 3). In successive iteration, each particle is evaluated by means of the fitness function specified in Eq. (11). Once the fitness of each particle in the swarm is obtained, its position and velocity are updated (step 9–19 of Algorithm 3). The entire procedure is repeated until a particular number of iteration is reached with the hope that a satisfactory solution will eventually be discovered. Once the specified number of iteration is finished, the particle i corresponding to maximum normalized overall membership value H^i is taken as the optimal fusion weights and the fused similarity score with these weights are taken as the ultimate retrieval result.

6 Combining re-ranking and rank aggregation methods for effective image retrieval

This section explores the feasibility of integrating re-ranking and rank aggregation methods to further improve the retrieval precision of CBIR systems. In the past, a lot of efforts have been made to devise more effective algorithms for image re-ranking and rank aggregation. However, none of them attempts to combine the advantages of these two approaches for better retrieval effectiveness. To this end, we formulate a novel image retrieval framework in which the proposed re-ranking and rank aggregation algorithms are efficiently integrated to yield better retrieval results.

In the proposed framework, the PSO-based rank list fusion scheme is used to combine the re-ranking results obtained with multiple image descriptors to form a single and more effective similarity matrix S_{opt} . The combination approach is formally defined as:

$$S_{opt} = \Phi(\Psi(S_1), \Psi(S_2), \dots, \Psi(S_n)) \quad (14)$$

where S_1, S_2, \dots, S_n , respectively, denotes the similarity matrices corresponding to n different types of image descriptors.

7 Performance evaluation and discussion

This section evaluates the retrieval efficiency of the proposed re-ranking and rank aggregation approaches and provides empirical evidences to demonstrate their superior performance over the traditional approaches. Moreover, the integration of the proposed re-ranking and rank aggregation strategies for the task of image retrieval is also evaluated. The rest of this section is organized as follows. A detailed description of the datasets used for evaluation is provided in Sect. 7.1. The quantitative indices used to measure the retrieval accuracy are described in Sect. 7.2. In Sect. 7.3, a brief description of the feature descriptors used in image retrieval experiments are provided. The experimental set-up

for evaluating the efficiency of the proposed re-ranking and rank aggregation schemes is outlined in Sect. 7.4. A comprehensive evaluation of the proposed re-ranking scheme is presented in Sect. 7.5. Section 7.6 validates the effectiveness of the proposed rank list fusion algorithm. The details of the statistical significance test conducted to assess the relevance of the proposed re-ranking and rank aggregation strategies is summarized in Sect. 7.7. Finally, the experimental analysis of the combination strategy for image retrieval is summarized in Sect. 7.8.

7.1 Description of the dataset

Four different datasets with contrasting properties are considered for evaluating the efficiency of the proposed dictionary learning scheme and the resulting image retrieval framework. The details of these four image collections are summarized below.

INRIA Holiday dataset [41] It involves 1491 high resolution images of different locations across the globe. The image collection is basically a mixture of natural scenes and man-made objects. Five hundred images in the collection are designated as queries and a predefined retrieval list is provided for each of these queries. An important characteristics of this dataset is that the images possesses high intra-class variance within each semantic concept. This property motivates us to select INRIA Holiday dataset as a benchmark to compare the efficiency of various image retrieval models.

Scene 15 dataset [63] This is mainly a collection of 4485 images grouped into 15 categories. The number of images per category varies from 210 to 410 and all the images have a fixed size of 300×250 pixels. There are mainly indoor and outdoor images in the collection. These images can be grouped into the following categories: bedroom (216 images), tall building (356 images), coast (360 images), city center (308 images), forest (328 images), highway (260 images) industrial (311 images), kitchen (210 images), living room (289 images), mountain (374 images), office (215 images), open country (410 images), store (315 images), street (292 images), suburb residences (241 images). This image collection serves as a good choice for evaluating the retrieval effectiveness of the proposed image re-ranking and rank list fusion schemes because it contains images with the same semantic concepts appearing in different contexts.

Oxford dataset [64] There are 5,062 building images of 11 various Oxford landmarks in this collection. Oxford dataset is widely acknowledged for its complexity to distinguish identical building facades from one another. Five images from each of the 11 landmarks are reserved as query and their corresponding retrieval lists are also provided as ground truth data. Thus, there are 55 queries to evaluate the proposed retrieval model. This dataset exhibits notable diversity among building images with variable appearances, positions, lighting conditions and view points. Hence, searching for similar images in response to a given query is highly challenging in this dataset.

Corel 10K dataset [65] There are 10000 images in Corel 10K dataset which spread over 100 concepts classes such

Table 1 Summary of various image descriptors used for evaluation

Descriptor type	Name of the descriptor	Acronym	Similarity measure used
Color	Weighted dominant color descriptor [50]	WDCD	DC-based similarity measure
	Compact color descriptor [51]	CCD	Color matching palette
	Pseudo-Zernike chromaticity distribution moment [42]	PZCDM	Euclidean distance
Texture	Steerable pyramid based texture feature [42]	SPTF	Euclidean distance
	Directional local extrema pattern [52]	DLEP	Canberra distance
	Local tetra patterns [53]	LTrP	d_1 distance
Shape	Angular radial transformation descriptor [54]	ARTD	Euclidean distance
	Generic Fourier descriptor [55]	GFD	Euclidean distance
	Curvature based Fourier descriptor [56]	CBFD	Euclidean distance
High-level descriptor	Bag-of-visual-words [57]	BoVW	HI-Kernel
	Vector of locally aggregated descriptor [58]	VLAD	Euclidean distance
	Object bank [59]	OB	L_1 -distance
	Sparse coding based Fisher vector coding [60]	SCFVC	Euclidean distance
	SParse output coding [61]	SPoC	Euclidean distance
	Sparse-NMF [62]	ℓ_0 - NMF	L_1 -distance

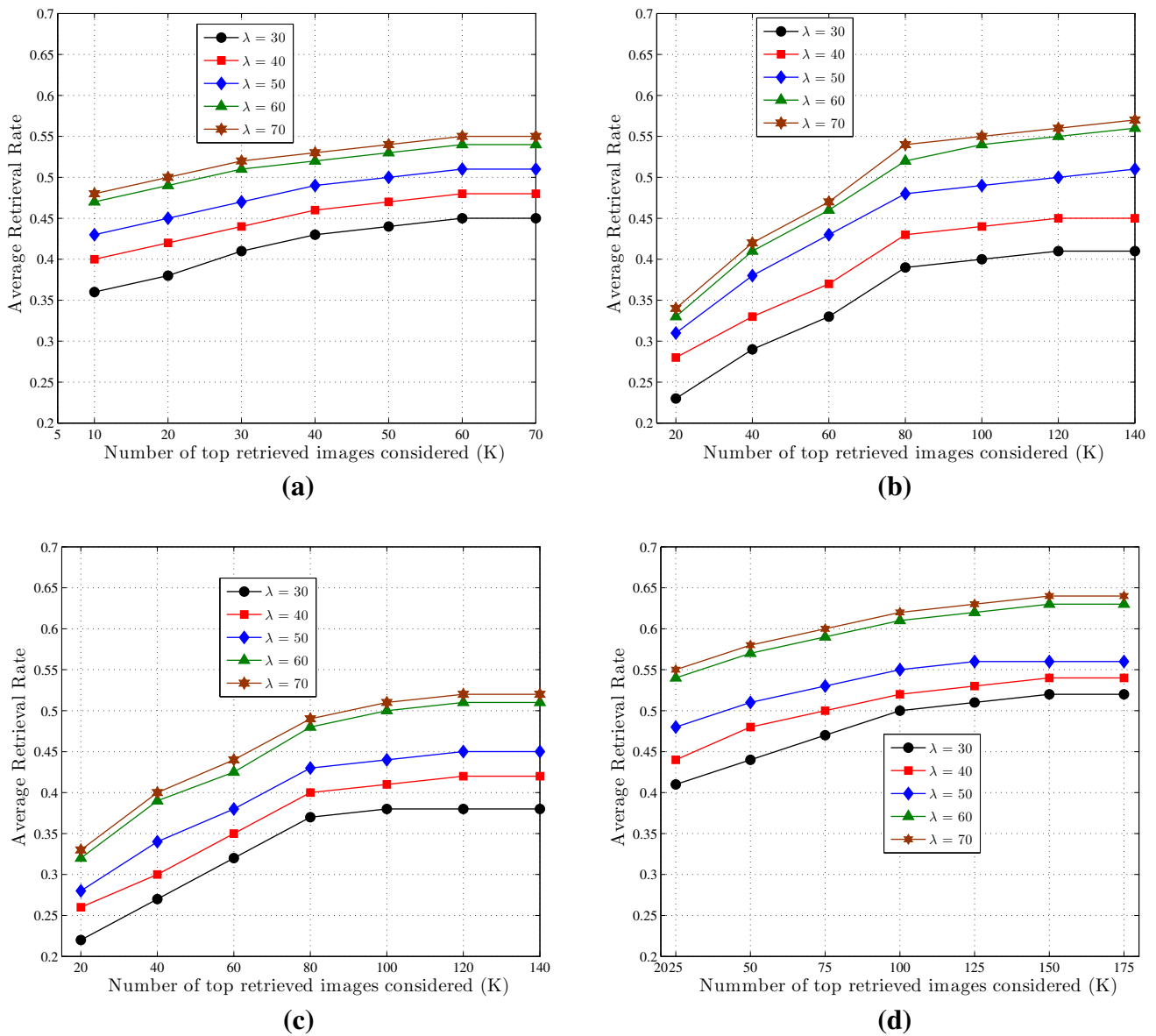


Fig. 3 Average retrieval rates obtained for various combinations of the parameters K and θ of the proposed image re-ranking scheme. **a** INRIA Holiday dataset. **b** Scene-15 dataset. **c** Oxford dataset. **d** Corel 10K dataset

as beach, flower, mountains, sunset etc. Each category contains 100 color images in JPEG format with a resolution of either 192×128 or 128×192 . A retrieved image is said to be relevant if and only if it is from the same category as that of the query. That is, any image selected from a test collection to act as a query will have exactly 99 relevant images in the collection. This dataset is quite challenging as it includes highly varying scene categories. As an example, images depicting the changes in color composition of “sky” viewed at regular time intervals during the day time is included in the dataset. Moreover, this dataset is enriched with sufficient number of images covering a diverse number of semantic concepts.

7.2 Evaluation metric

An image retrieval system generates a ranked list of images belonging to a particular dataset in response to a submitted query. The rank of an image is determined by its relevance to the query at hand. To be able to compare various image retrieval models, first a set of performance measures are to be identified. When the ground truth of the dataset is available, the system’s performance is generally measured in terms of quantitative metrics such as precision and recall. The precision of a retrieval system measures the percentage of relevant images in the ranked retrieval list and the recall denotes the percentage of relevant images retrieved by the system. These two metrics are defined as follows:

Table 2 Re-ranking results of the proposed scheme for low-level descriptors

Descriptor type	Descriptor name	MAP			P@20			Avg R-precision		
		Before re-ranking	After re-ranking	Gain (%)	Before re-ranking	After re-ranking	Gain (%)	Before re-ranking	After re-ranking	Gain (%)
INRIA Holiday dataset										
Color	CCD	0.2735	0.3886	42.08	0.2925	0.4168	42.49	0.2912	0.3982	36.74
	WDCD	0.3495	0.4615	32.04	0.3588	0.4584	27.75	0.3327	0.4774	43.49
Texture	PZCDM	0.4325	0.5638	30.35	0.4566	0.5842	27.94	0.4413	0.5767	30.68
	DLEP	0.2964	0.4281	44.43	0.3142	0.4349	38.41	0.3110	0.4284	37.74
Shape	LTrP	0.3593	0.4766	32.64	0.3661	0.4718	28.87	0.3457	0.4648	34.45
	SPTF	0.4471	0.5754	28.69	0.4584	0.5864	27.92	0.4372	0.5683	29.98
Scene 15 dataset	ARTD	0.2287	0.3688	61.25	0.2483	0.3637	46.47	0.2289	0.3525	53.99
	GFD	0.2649	0.3729	40.77	0.2992	0.3928	35.822	0.2732	0.3977	45.57
Color	CBFD	0.3014	0.4139	37.32	0.3153	0.4178	32.50	0.2985	0.4168	39.63
	CCD	0.3535	0.4882	38.10	0.3712	0.4928	32.75	0.3366	0.4569	35.73
Texture	WDCD	0.4086	0.5229	27.97	0.4176	0.5368	28.54	0.4002	0.5286	32.08
	PZCDM	0.4564	0.5806	27.21	0.4611	0.5739	24.46	0.4479	0.5718	27.66
Shape	DLEP	0.3484	0.4773	36.99	0.3684	0.4852	31.70	0.3449	0.4782	38.64
	LTrP	0.4148	0.5286	27.43	0.4218	0.5327	26.29	0.3996	0.5185	29.75
Oxford dataset	SPTF	0.4662	0.5862	25.74	0.4807	0.5962	24.02	0.4734	0.5828	23.10
	ARTD	0.2445	0.3624	48.22	0.2678	0.3749	39.99	0.2531	0.3684	45.55
Color	GFD	0.2777	0.3867	39.25	0.2940	0.4131	40.51	0.2811	0.4074	44.93
	CBFD	0.3159	0.4294	35.92	0.3316	0.4328	30.51	0.3194	0.4249	33.03
Texture	CCD	0.2866	0.3796	32.44	0.2990	0.3877	29.66	0.2811	0.3768	34.04
	WDCD	0.3472	0.4688	35.02	0.3674	0.4791	30.40	0.3355	0.4586	36.69
Shape	PZCDM	0.4020	0.5122	27.41	0.4226	0.5464	29.29	0.3949	0.5171	30.94
	DLEP	0.3154	0.4379	38.83	0.3353	0.4452	32.77	0.3211	0.4582	42.69
Oxford dataset	LTrP	0.3922	0.5252	33.91	0.4168	0.5383	29.15	0.4026	0.5183	28.73
	SPTF	0.4297	0.5432	26.41	0.4421	0.5619	27.09	0.4388	0.5613	27.91
Color	ARTD	0.3041	0.4153	36.56	0.3162	0.4376	38.39	0.2919	0.4274	46.42
	GFD	0.3356	0.4547	35.48	0.3497	0.4747	35.74	0.3286	0.4542	38.22
Texture	CBFD	0.3768	0.4986	32.32	0.3973	0.5068	27.56	0.3731	0.4872	30.58

Table 2 continued

Descriptor type	Descriptor name	MAP		P@20		Avg R-precision			
		Before re-ranking	After re-ranking	Gain (%)	Before re-ranking	After re-ranking	Gain (%)		
Corel-10K dataset	Color	CCD	0.3148	0.3959	0.3313	0.4234	0.3176	0.4071	28.18
		WDGD	0.3696	0.4777	0.3794	0.4868	0.3524	0.4783	35.72
		PZCDM	0.3909	0.5212	0.4122	0.5327	0.3916	0.5128	30.94
Texture	DLEP	0.2957	0.3884	0.3142	0.3972	0.3025	0.3975	31.40	
	LTrP	0.4118	0.5337	0.4363	0.5689	0.4182	0.5236	25.20	
Shape	SPTF	0.4369	0.5667	0.4482	0.5679	0.4302	0.5793	34.65	
	ARTD	0.2869	0.4089	0.3014	0.4281	0.2957	0.4172	41.08	
	GFD	0.3210	0.4581	0.3461	0.4717	0.3352	0.4547	35.65	
	CBFD	0.3554	0.4751	0.3688	0.4857	0.3428	0.4579	33.57	

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (15)$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the set}} \quad (16)$$

Precision and recall do not take into account the order in which relevant images appear in the ranked retrieval list. When two retrieval systems have the same precision and recall values, the system that ranks relevant images higher is mostly preferred. In order to solve this issue, measures like Precision at k ($P@k$) and R-precision are introduced. $P@k$ is the value of precision calculated using the first k documents in the retrieval list. Similarly, R-precision for a given query is defined to be the precision after retrieving R images from the image data base and is expressed as:

$$R - \text{Precision} = \frac{1}{R} \sum_{j=1}^R \text{Rel}(j) \quad (17)$$

where R is the total number of relevant images in the database for the given query and $\text{Rel}(j)$ is an indicator function which returns the value 1 when the image present at the j -th location of the retrieval list is relevant with respect to the given query.

Moreover, precision can be expressed as a function of recall. The interpolated precision recall graph plots precision as a function of recall and can be used to assess the overall performance of the retrieval framework. The interpolated precision p_{int} at a recall level r_i is calculated as the largest observed precision for any recall value r between r_i and r_{i+1} :

$$P_{\text{int}}(r_i) = \max_{r_i \leq r \leq r_{i+1}} \text{Precision}(r) \quad (18)$$

An alternative single valued evaluation metric is the mean average precision (MAP) and is defined as:

$$\text{Mean Average Precision (mAP)} = \frac{1}{|Q|} \sum_{q \in Q} \text{AP}(q) \quad (19)$$

where $|Q|$ denotes the number of images in the query set Q , $\text{AP}(q)$ is the average precision for a given query $q \in Q$ and is defined as the ratio of the sum of precision values from rank positions where a relevant image is found in the retrieval result to the total number of relevant images in the database.

One last metric is the Average Retrieval Rate (ARR) and is defined as:

$$\text{Average Retrieval Rate (ARR)} = \frac{1}{N_Q} \sum_{q=1}^{N_Q} \text{RR}(q) \quad (20)$$

where N_Q represents the number of queries used for evaluating the retrieval system. $\text{RR}(q)$ is the retrieval rate for a single query q and is calculated as:

Table 3 Re-ranking results of the proposed scheme for high-level descriptors

Descriptor name	MAP			P@20			Avg R-Precision		
	Before re-ranking	After re-ranking	Gain (%)	Before re-ranking	After re-ranking	Gain (%)	Before re-ranking	After re-ranking	Gain (%)
INRIA Holiday dataset									
BoVW	0.5278	0.6343	20.17	0.5564	0.6643	19.39	0.5132	0.6363	23.98
VLAD	0.5543	0.6781	22.33	0.5894	0.6986	18.52	0.5398	0.6629	22.80
OB	0.5732	0.6873	19.90	0.5961	0.7059	18.41	0.5679	0.6761	19.05
SCFVC	0.6081	0.7153	17.62	0.6153	0.7349	19.43	0.5811	0.7227	24.36
SPoC	0.6314	0.7436	17.77	0.6541	0.7951	21.55	0.6321	0.7775	23.00
ℓ_0 -NMF	0.6480	0.7719	19.12	0.6603	0.8068	22.18	0.6551	0.7961	21.52
Scene-15 dataset									
BoVW	0.5169	0.6437	24.53	0.5368	0.6583	22.63	0.5128	0.6349	23.81
VLAD	0.5371	0.6473	20.51	0.5529	0.6774	22.51	0.5281	0.6477	22.64
OB	0.5544	0.6727	21.33	0.5681	0.6828	20.19	0.5463	0.6628	21.32
SCFVC	0.5736	0.7065	23.16	0.5919	0.7266	24.86	0.5782	0.7082	22.48
SPoC	0.6172	0.7383	19.62	0.6277	0.7576	20.69	0.6057	0.7384	21.90
ℓ_0 -NMF	0.6430	0.7662	19.16	0.6676	0.7948	19.05	0.6442	0.7641	18.61
Oxford dataset									
BoVW	0.5015	0.6181	23.25	0.5161	0.6262	21.33	0.4959	0.6039	21.77
VLAD	0.5179	0.6228	20.25	0.5286	0.6336	19.86	0.5066	0.6382	25.97
OB	0.5271	0.6363	20.71	0.5393	0.6548	21.41	0.5141	0.6472	25.88
SCFVC	0.5422	0.6667	22.96	0.5514	0.6885	24.86	0.5017	0.6337	26.31
SPoC	0.5639	0.6883	22.06	0.5778	0.7081	22.55	0.5497	0.6771	23.17
ℓ_0 -NMF	0.5930	0.7275	22.68	0.6069	0.7447	22.70	0.5889	0.7219	22.58
Corel-10K dataset									
BoVW	0.5273	0.6458	22.47	0.5474	0.6563	1.89	0.5161	0.6269	21.46
VLAD	0.5484	0.6645	21.17	0.5693	0.6874	20.74	0.5410	0.6582	21.66
OB	0.5684	0.6881	21.05	0.5747	0.6907	20.18	0.5520	0.6759	22.44
SCFVC	0.5872	0.7129	21.40	0.5973	0.7258	21.51	0.6027	0.7269	20.60
SPoC	0.6074	0.7362	21.20	0.6227	0.7574	21.63	0.6094	0.7328	20.24
ℓ_0 -NMF	0.6279	0.7717	22.90	0.6373	0.7718	21.10	0.6116	0.7539	23.26

$$RR(q) = \frac{N_R(\alpha, q)}{N_G(q)} \quad (21)$$

where $N_G(q)$ is the number of ground truth images of a query q and $N_R(\alpha, q)$ indicates the number of relevant images found in the first $\alpha \times N_G(q)$ images. The value of α should be greater than or equal to 1. Selecting larger α values would be less discriminative between very good retrieval results and those results that are not so good ones. Therefore, this work set the value of α as 1.5.

In addition to this standard measures, the effectiveness of image re-ranking algorithms is evaluated by means of relative gain. Let rs_b be the value of the retrieval score before the use of the re-ranking or rank aggregation algorithm and let rs_a be the value after applying the proposed scheme, then relative gain is computed as follows:

$$\text{Relative Gain} = \frac{rs_a - rs_b}{rs_b} \quad (22)$$

7.3 Image descriptors used for retrieval experiments

CBIR systems perform image retrieval on the basis of feature vectors automatically extracted from image pixels. In general, feature vectors are meaningful abstraction of image data and it usually encode the visual contents of images in a compact fashion. The visual contents of images can be either low-level or high-level (semantic). In general, the most effective low-level features used for representing the visual contents of images are color, texture and shape. On the other hand, the high-level content is the actual meaning captured by humans when they look at the images. Thus, the feature vector must encode the visual properties of an image in such a way that allows it to be compared and matched with other

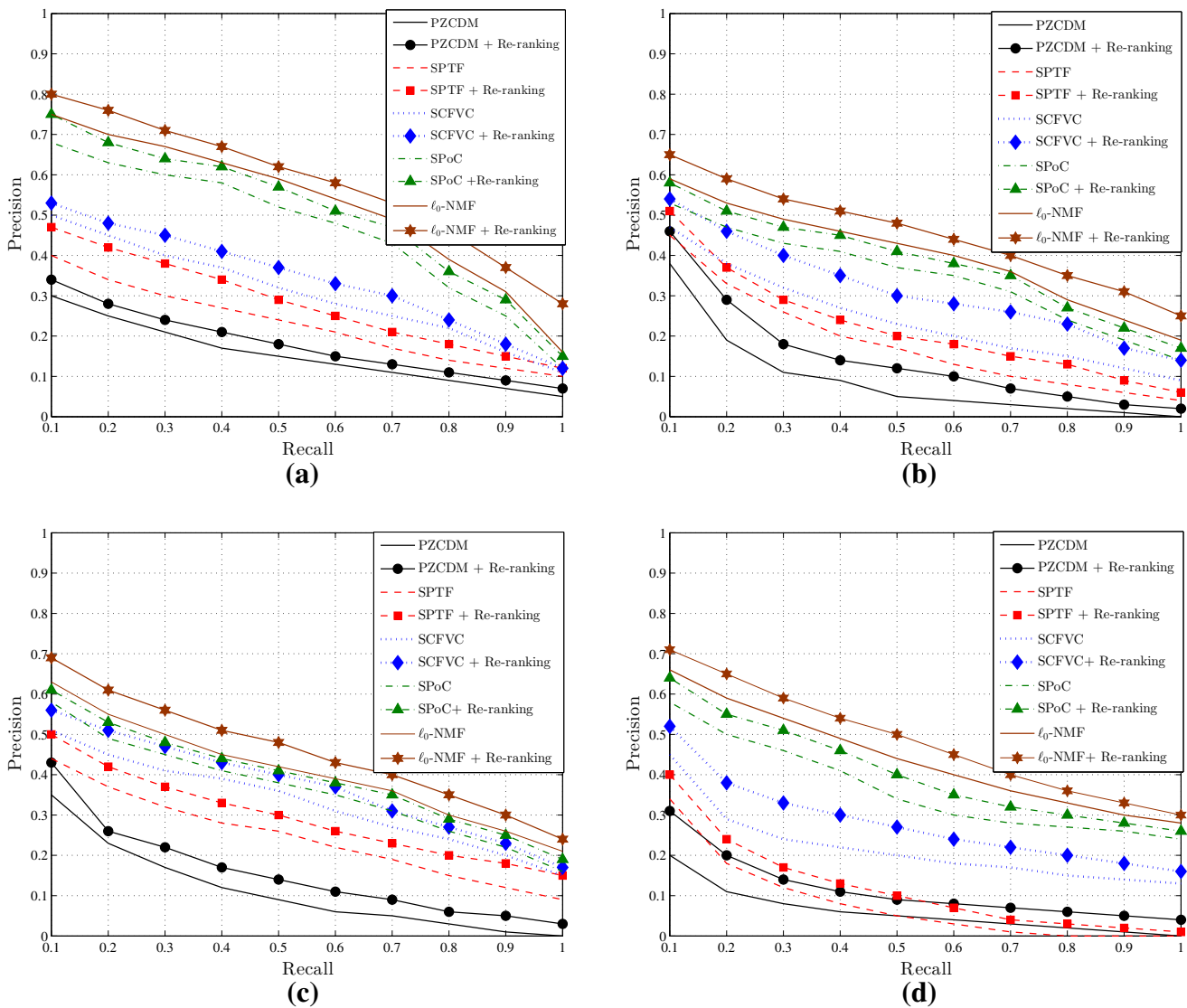


Fig. 4 Retrieval performance of the proposed image re-ranking scheme based on 11-point interpolated average precision. **a** INRIA Holiday dataset. **b** Scene-15 dataset. **c** Oxford data set. **d** Corel 10K dataset

images in the collection to find a perfect match. Distance functions are the most simple and the widely used metric to judge the similarity among feature vectors. By using distance functions, the retrieval system generates an ordered retrieval list in increasing order of the distance calculated among the feature vectors derived from database images and the given query image.

However, deriving a universal descriptor that gives high retrieval precision for all sorts of datasets is still an open problem in image retrieval domain. Each of the descriptor, whether it is low-level or high-level, has its own merit and demerit. Moreover, those descriptors that belongs to the same category are always complementary in nature. In this paper, we make use of low-level as well as high-level descriptors to assess the effectiveness of the proposed

post-retrieval optimization framework. Therefore, a set of representative candidates that provides state-of-the-art performance in image retrieval have been selected from each of the above-mentioned descriptor categories to evaluate the proposed image re-ranking and rank aggregation schemes. The set of all image descriptors selected for evaluation together with the corresponding similarity measures used in the retrieval experiments are summarized in Table 1.

7.4 Experimental protocol

The retrieval experiments using high-level descriptors are carried out using tenfold cross validation. To do so, images in the database are arbitrarily split into ten folds roughly of the same size. In each experiment, nine image subsets are used

Table 4 Comparative evaluation of various image re-ranking schemes

Algorithm used	Descriptors used	MAP	P@20	Avg R-precision	Algorithm used	Descriptors used	MAP	P@20	Avg R-precision
INRIA Holiday dataset									
DOA	PZCDM	0.4696	0.4941	0.4853	Scene 15 dataset	PZCDM	0.4922	0.4982	0.4852
	SPTF	0.4831	0.4986	0.4726		SPTF	0.5046	0.5171	0.5117
RL-Sim	SCFVC	0.6357	0.6521	0.6234	RL-Sim	SCFVC	0.6274	0.6378	0.6358
	SPoC	0.6613	0.6936	0.6728		SPoC	0.6452	0.6653	0.6479
	ℓ_0 -NMF	0.6772	0.7019	0.6925		ℓ_0 -NMF	0.6658	0.6748	0.6571
	PZCDM	0.4879	0.5152	0.5098		PZCDM	0.5169	0.5157	0.5076
RKNN-ML	SPTF	0.5014	0.5139	0.4964	RKNN-ML	SPTF	0.5288	0.5336	0.5338
	SCFVC	0.6578	0.6715	0.6479		SCFVC	0.6487	0.6546	0.6534
	SPoC	0.6761	0.7162	0.6948		SPoC	0.6639	0.6884	0.6629
	ℓ_0 -NMF	0.6884	0.7248	0.7132		ℓ_0 -NMF	0.6851	0.6933	0.6726
Proposed Algorithm	PZCDM	0.5068	0.5386	0.5373	Proposed Algorithm	PZCDM	0.5363	0.5390	0.5336
	SPTF	0.5274	0.5360	0.5285		SPTF	0.5476	0.5575	0.5582
	SCFVC	0.6781	0.7474	0.6793		SCFVC	0.6694	0.6787	0.6834
	SPoC	0.6969	0.7368	0.7269		SPoC	0.6921	0.7074	0.6922
Oxford dataset	NMF	0.7131	0.7501	0.7487	Corel 10K dataset	NMF	0.7127	0.7279	0.7057
	PZCDM	0.5638	0.5842	0.5767		PZCDM	0.5806	0.5739	0.5718
	SPTF	0.5754	0.5864	0.5683		SPTF	0.5862	0.5962	0.5828
	SCFVC	0.7193	0.7349	0.7227		SCFVC	0.7065	0.7266	0.7082
DOA	SPoC	0.7436	0.7951	0.7775	DOA	SPoC	0.7383	0.7556	0.7384
	ℓ_0 -NMF	0.7719	0.8068	0.7961		ℓ_0 -NMF	0.7662	0.7948	0.7641
	PZCDM	0.4456	0.4684	0.4376		PZCDM	0.4356	0.4584	0.4347
	SPTF	0.4673	0.4859	0.4775		SPTF	0.4738	0.4826	0.4739
RL-Sim	SCFVC	0.5881	0.5963	0.5466	RL-Sim	SCFVC	0.6223	0.6357	0.6461
	SPoC	0.6035	0.6196	0.5848		SPoC	0.6437	0.6639	0.6485
	ℓ_0 -NMF	0.6342	0.6437	0.6279		ℓ_0 -NMF	0.6639	0.6715	0.6579
	PZCDM	0.4562	0.4827	0.4584		PZCDM	0.4526	0.4774	0.4591
RKNN-ML	SPTF	0.4892	0.5092	0.4986	RKNN-ML	SPTF	0.4986	0.5021	0.4949
	SCFVC	0.6014	0.6135	0.5681		SCFVC	0.6408	0.6517	0.6688
	SPoC	0.6271	0.6341	0.6033		SPoC	0.6674	0.6829	0.6653
	ℓ_0 -NMF	0.6573	0.6626	0.6476		ℓ_0 -NMF	0.6849	0.6918	0.6748
SPTF	PZCDM	0.4832	0.5109	0.4868	SPTF	PZCDM	0.4897	0.5084	0.4893
	SPTF	0.5116	0.5311	0.5287		SPTF	0.5201	0.5336	0.5224

Table 4 continued

Algorithm used	Descriptors used	MAP	P@20	Avg R-precision	Algorithm used	Descriptors used	MAP	P@20	Avg R- Precision
Proposed Algorithm	SCFVC	0.6383	0.6490	0.5927	Proposed Algorithm	SCFVC	0.6737	0.6819	0.6965
	SPoC	0.6569	0.6679	0.6337		SPoC	0.6949	0.7122	0.6896
	ℓ_0 -NMF	0.6888	0.6927	0.6731		ℓ_0 -NMF	0.7128	0.7203	0.7016
	PZCDM	0.5122	0.5464	0.5171		PZCDM	0.5212	0.5327	0.5128
Proposed Algorithm	SPTF	0.5432	0.5619	0.5613	Proposed Algorithm	SPTF	0.5667	0.5679	0.5793
	SCFVC	0.6667	0.6885	0.6337		SCFVC	0.7129	0.7258	0.7269
	SPoC	0.6983	0.7081	0.6771		SPoC	0.7362	0.7574	0.7328
	ℓ_0 -NMF	0.7375	0.7447	0.7219		ℓ_0 -NMF	0.7717	0.7718	0.7539

for training the model and the remaining subset will function as the query. Hence, each image subset is used once as the query. The evaluation metrics are then computed as the average over these ten trials. For low-level features, the evaluation metrics are calculated as the mean value by considering all the database images as the query. All the experiments are carried out in MATLAB 2013b on an Intel Core i7-3770, 3.40 GHz desktop PC equipped with 16 GB of RAM and Ubuntu 64 bit operating system.

7.5 Evaluation of the proposed image re-ranking scheme

This section illustrates the retrieval results of the proposed image re-ranking scheme. Section 7.5.1 analyzes the impact of various parameters of the distance correlation coefficient-based image re-ranking algorithm on the retrieval effectiveness. Section 7.5.2 provides a comparative evaluation of the proposed re-ranking algorithm.

7.5.1 Impact of parameters

The parameters of the proposed image re-ranking algorithm have great implication on the final retrieval performance. Therefore, determining optimal values for these parameters is a challenging task. The proposed distance correlation-based image re-ranking algorithm mainly depends on two parameters: (i) K number of top retrieved images (ii) θ the threshold for the distance correlation measure. Optimal values for these parameters are estimated in terms of average retrieval rates. For all the datasets considered for evaluation, the average retrieval rates are computed with K values ranging from 20 to 200 and five different threshold (θ) values {30, 40, 50, 60, 70}. The average retrieval rate obtained for various datasets while changing K along with θ is depicted in Fig. 3. From these results, it can be concluded that for small values of K and θ , the retrieval system fails to yield acceptable precision. As the threshold (θ) increases, then even for small values of K the retrieval system can achieve better retrieval precision. Considering all these factors into account, the number of top retrieved images considered for evaluation (K) and the threshold (θ) are fixed to 100 and 70.

7.5.2 Retrieval results

In this section, the set of experiments conducted for demonstrating the effectiveness of the proposed image re-ranking scheme is presented. Various image re-ranking schemes such as Distance Optimization Algorithm [33], RL-Sim re-ranking algorithm [15] and Reciprocal kNN Graphs based manifold learning (RKNN-ML) algorithm [34] have been evaluated in comparison with the proposed scheme by considering both low-level and high-level descriptors for all the four datasets.

Table 5 Comparative evaluation of various metaheuristics algorithm for the task of rank list fusion

Method used	Best H^i value	No of iterations	Best H^i value	No of iterations	
		INRIA Holiday dataset		Scene 15 dataset	
GA	0.746	1200	0.765	1325	
DE	0.856	800	0.875	975	
ABC	0.787	1050	0.806	1200	
CSA	0.812	950	0.838	1050	
PSO	0.889	610	0.907	750	
		Oxford dataset		Corel 10K dataset	
GA	0.758	1285	0.771	1450	
DE	0.866	900	0.884	1050	
ABC	0.798	1100	0.813	1375	
CSA	0.824	1000	0.836	1225	
PSO	0.897	690	0.917	925	

Table 6 Rank aggregation results of the proposed scheme for low-level descriptors

Descriptor type	Descriptor used	MAP	P@20	Avg R-precision	MAP	P@20	Avg R-precision
		Holiday dataset			Scene-15 dataset		
Color	WDCD	0.3495	0.3588	0.3327	0.4086	0.4176	0.3802
	PZCDM	0.4325	0.4566	0.4413	0.4564	0.4611	0.4279
	WDCD + PZCDM	0.5482	0.5623	0.5571	0.5712	0.5803	0.5538
Texture	LTrP	0.3593	0.3661	0.3457	0.4148	0.4218	0.3996
	SPTF	0.4471	0.4584	0.4372	0.4662	0.4807	0.4734
	LTrP + SPTF	0.5529	0.5617	0.5643	0.5778	0.5914	0.5862
Shape	GFD	0.2649	0.2892	0.2732	0.2777	0.2940	0.2811
	CBFD	0.3014	0.3153	0.2985	0.3159	0.3316	0.3194
	GFD + CBFD	0.4250	0.4349	0.4037	0.4325	0.4516	0.4222
		Oxford dataset			Corel 10K dataset		
Color	WDCD	0.3472	0.3674	0.3355	0.3696	0.3794	0.3524
	PZCDM	0.4020	0.4226	0.3949	0.3909	0.4122	0.3916
	WDCD + PZCDM	0.5262	0.5439	0.5314	0.5094	0.5275	0.5199
Texture	LTrP	0.3922	0.4168	0.4026	0.4118	0.4363	0.4182
	SPTF	0.4297	0.4421	0.4388	0.4369	0.4482	0.4302
	LTrP + SPTF	0.5387	0.5561	0.5438	0.5419	0.5596	0.5416
Shape	GFD	0.3356	0.3497	0.3286	0.3210	0.3461	0.3352
	CBFD	0.3768	0.3973	0.3731	0.3554	0.3688	0.3428
	GFD + CBFD	0.4852	0.5047	0.3883	0.4664	0.4758	0.4015

Tables 2 and 3 summarizes the mean average precision, P@20 and average R-precision values obtained with the proposed approach for various low-level and high-level descriptors under the following circumstances: before and after the use of the proposed re-ranking scheme in image retrieval. For each of the above-mentioned evaluation metrics, the relative gain achieved with the proposed model is also reported. All these results shows that the distance correlation-based re-ranking scheme is more effective in image retrieval and there is significant gain in the retrieval

performance as compared to the results of individual descriptors in isolation.

Figure 4 depicts the 11-point interpolated average precision curve obtained while employing a selected set of low-level and high-level descriptors in isolation and in combination with the proposed re-ranking scheme for the retrieval experiments. Across all values of recall, it seems that the retrieval precision achieved by the proposed re-ranking scheme is significantly better than individual descriptors used in isolation for all image collections.

Table 7 Rank aggregation results of the proposed scheme for high-level descriptors

Descriptor name	MAP	P@20	Avg R-precision	MAP	P@20	Avg R-precision
	Holiday dataset			Scene-15 dataset		
SCFVC	0.6081	0.6153	0.5811	0.5736	0.5819	0.5782
SPoC	0.6314	0.6541	0.6321	0.6172	0.6277	0.6057
ℓ_0 -NMF	0.6480	0.6603	0.6551	0.6430	0.6676	0.6442
SCFVC+ SPoC + ℓ_0 -NMF	0.7992	0.8156	0.8067	0.7943	0.8128	0.7994
	Oxford dataset			Corel 10K dataset		
SCFVC	0.5422	0.5514	0.5017	0.5872	0.5973	0.6027
SPoC	0.5639	0.5778	0.5497	0.6074	0.6227	0.6094
ℓ_0 -NMF	0.5930	0.6069	0.5889	0.6279	0.6373	0.6116
SCFVC+ SPoC + ℓ_0 -NMF	0.7485	0.7352	0.7384	0.7791	0.7815	0.7638

The comparative evaluation of the proposed image re-ranking scheme is outlined in Table 4. It can be observed that the distance correlation-based image re-ranking scheme accomplished significant gain in retrieval effectiveness in case of all four datasets and all types of image descriptors as compared to other existing methods. On an average, the proposed re-ranking model achieved 6% improvement in overall retrieval effectiveness across all the four dataset considered for evaluation. These results underline the fact that the proposed image re-ranking scheme yields favorable retrieval scores in comparison with state-of-the-art approaches.

7.6 Evaluation of the PSO-based rank list fusion scheme

A detailed evaluation of the proposed PSO-based rank list fusion scheme is presented in this section. The procedure used for similarity score normalization and the retrieval experiments carried out in various datasets using the proposed rank list fusion scheme are comprehensively discussed in the rest of the subsections.

7.6.1 Similarity matrix normalization

It should be noted that the physical meaning of individual feature descriptors are different and the corresponding similarity matrices need not be on the same numerical scale. That is, the similarity matrices at the output of individual retrieval models may not be homogeneous. Therefore, these similarity matrices cannot be directly aggregated and normalization has to perform before the actual fusion takes place. The scaling down transformation of the original similarity matrix to a reasonably lower range is termed as normalization. As it is a critical step in similarity score fusion, the normalization process must be carefully designed.

The tanh-estimator introduced by Hampel et al. [66] is reported to be an efficient and robust normalization technique. This paper adopts tanh-estimators for the normal-

ization process. Let $\{S_i\}_{i=1}^N$ be the set of similarity matrices of N database images corresponding to a given query image I_q and $\{\mu, \sigma\}$ be the mean and the standard deviation estimates of these similarity scores. Then, for each image in the database, the normalized similarity score based on tanh-estimator is given by:

$$\hat{S}_i = \frac{1}{2} \left\{ \tanh \left(0.01 * \left(\frac{S_i - \mu}{\sigma} \right) \right) + 1 \right\} \quad (23)$$

where S_i is the original similarity score and \hat{S}_i is its normalized version after applying tanh-estimator.

7.6.2 Retrieval result

We consider the reciprocal rank fusion strategy (RRF) [9], the distance optimization algorithm based clustering (DOA-Cluster) [33], and the query-specific rank fusion algorithm (QSRF) [27] as the baseline to evaluate the proposed rank list fusion scheme. Based on the retrieval results of the image re-ranking experiments presented in Tables 2 and 3, the best four among low-level descriptors and the best three among high-level descriptors are selected for the task of rank list fusion. Thus, PZCDM [42], SPTF [42], WDCD [50] and LTrP [53] are selected from the category of low-level descriptors and SCFVC [60], SPoC [61] and ℓ_0 -NMF [62] are chosen from the family of high-level descriptors.

First of all, Table 5 summarizes the result obtained by the proposed PSO-based approach in solving the optimization problem specified in Eq. (11) in comparison with other meta heuristic algorithms. The table provides the best H^i value obtained by each of the approaches and the number of iterations performed by each methods to reach the corresponding best H^i values. In all the datasets selected for evaluation, the proposed approach converged to a better H^i values in lesser number of iterations. Thus, it can be conclude that proposed PSO-based approach is better than other metaheuristic algorithms such as GA [44], DE

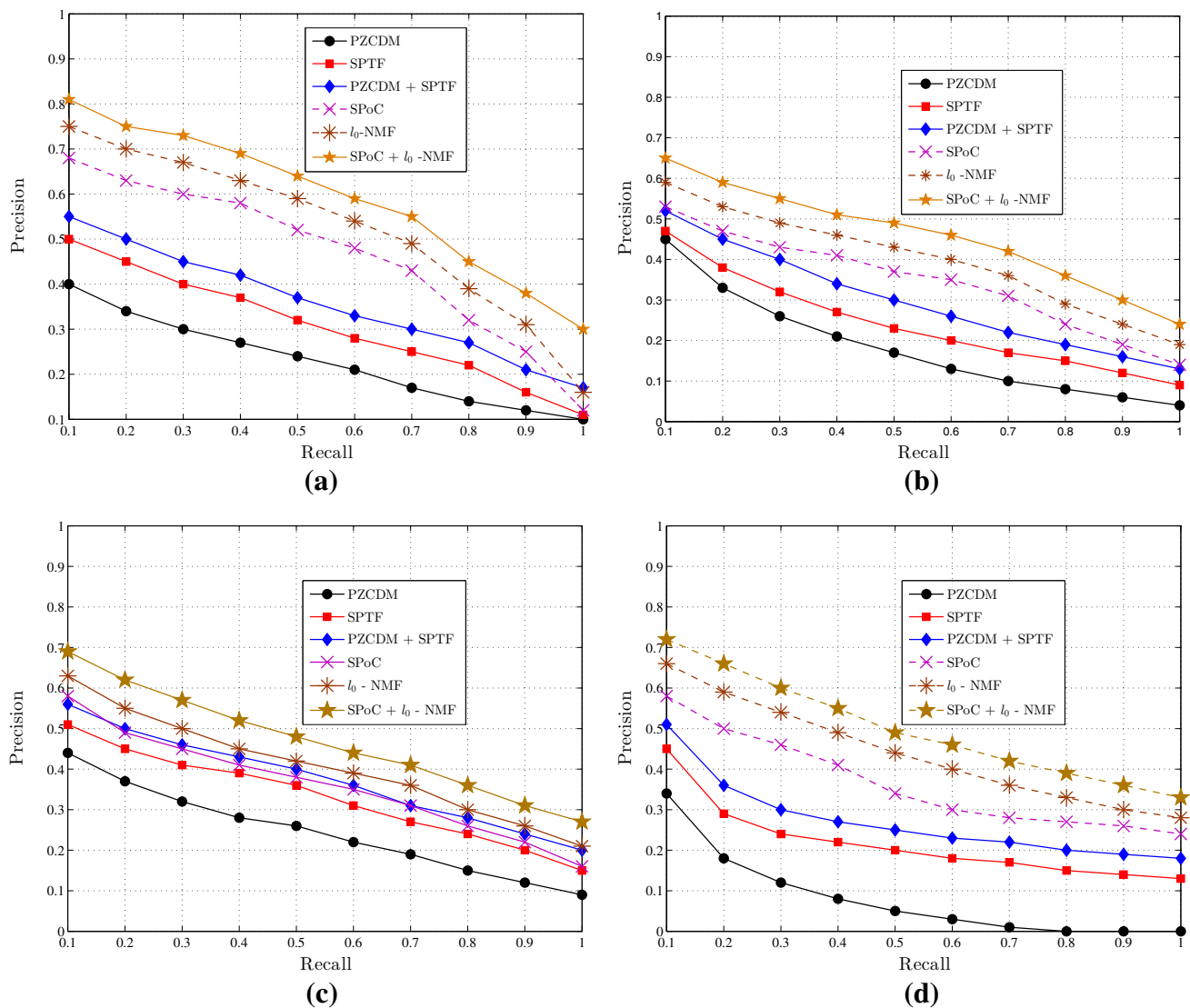


Fig. 5 Retrieval performance of the proposed rank aggregation scheme based on 11-point interpolated average precision. **a** INRIA Holiday dataset. **b** Scene-15 dataset. **c** Oxford data set. **d** Corel 10K dataset

[46], ABC [47] and CSA [48] for the task of rank list fusion.

Then, we analyze the relative improvements in MAP, P@20 and Avg. R-precision values while using the PSO-based rank list fusion scheme for image retrieval operation based on the results summarized in Tables 6 and 7. In contrast to the retrieval result of each descriptor in isolation, there is significant gain in precision while using PSO-based rank list fusion scheme. For example, in INRIA Holiday dataset [41], the MAP score is increased to 79.92% by using the PSO-based approach for aggregating the retrieval lists returned by SCFVC, SPoC and l_0 -NMF descriptors. When these descriptors are used in isolation for the retrieval task, respective gains of 60.81, 63.14 and 64.80% are obtained. Similar results can also be observed with P@20 and Average R-precision values.

Figure 5 shows the 11-point interpolated precision curves of certain selected image descriptors in different situations: i.e., before rank list fusion and after applying the PSO-based rank list fusion algorithm. It can be easily inferred that the precision achieved by the proposed rank list fusion scheme is notably higher than that of individual descriptors in isolation.

Next, a comparative evaluation of the PSO-based rank list fusion scheme is provided. Table 8 summarizes the mean average precision (MAP), Precision at 20 (P@20) and Average R-precision values obtained with the proposed PSO-based rank list fusion scheme in comparison with state-of-the-art approaches. It is well understood from the above results that for all combination of image descriptors, we can observe positive gain in retrieval precision as compared to the state-of-the-art approaches. While analyzing

Table 8 Comparative evaluation of various rank aggregation schemes

Algorithm used	Descriptors used	MAP	P@20	Avg R-precision	Algorithm used	Descriptors used	MAP	P@20	Avg R-precision
INRIA Holiday dataset									
RRF	WDCD + LTpP	0.3863	0.3981	0.3749	RRF	WDCD + LTpP	0.4439	0.4538	0.4264
	PZCDM + SPTF	0.4733	0.4879	0.4782		PZCDM + SPTF	0.4983	0.5127	0.5039
DOA-Cluster	SCFVC+ SPoC	0.6649	0.6853	0.66665	DOA-Cluster	SCFVC+ SPoC	0.6449	0.6592	0.6384
	SpoC + ℓ_0 -NMF	0.6758	0.6984	0.6895		SpoC +NMF	0.6773	0.6963	0.6738
	SCFVC+ SpoC+ ℓ_0 -NMF	0.6992	0.7156	0.7067		SCFVC+ SpoC+ ℓ_0 -NMF	0.6943	0.7128	0.6991
	WDCD + LTpP	0.4172	0.4293	0.4062		WDCD + LTpP	0.4746	0.4856	0.4571
	PZCDM + SPTF	0.5047	0.5186	0.5097		PZCDM + SPTF	0.5239	0.65442	0.5352
QSRF	SCFVC+ SPoC	0.6953	0.7161	0.6970	QSRF	SCFVC+ SPoC	0.6758	0.6884	0.6691
	SpoC + ℓ_0 -NMF	0.7053	0.7263	0.7188		SpoC + ℓ_0 -NMF	0.7080	0.7276	0.7053
	SCFVC+ SpoC + ℓ_0 -NMF	0.7284	0.74638	0.7377		SCFVC+ SpoC+ ℓ_0 -NMF	0.7251	0.7447	0.7283
	WDCD + LTpP	0.4355	0.4474	0.4238		WDCD + LTpP	0.4926	0.5029	0.4759
	PZCDM + SPTF	0.5227	0.5365	0.5274		PZCDM + SPTF	0.5426	0.5619	0.5523
Proposed Algorithm	SCFVC+ SPoC	0.7137	0.7346	0.7149	Proposed algorithm	SCFVC+ SPoC	0.6938	0.7081	0.6876
	SpoC + ℓ_0 -NMF	0.7238	0.7436	0.7384		SpoC + ℓ_0 -NMF	0.7261	0.7455	0.7243
	SCFVC+ SpoC+ ℓ_0 -NMF	0.7483	0.7642	0.7556		SCFVC+ SpoC+ ℓ_0 -NMF	0.7436	0.7617	0.7483
	WDCD + LTpP	0.4863	0.4981	0.4749		WDCD + LTpP	0.5439	0.5538	0.5264
	PZCDM + SPTF	0.5733	0.5879	0.5782		PZCDM + SPTF	0.5939	0.6128	0.6039
Oxford dataset	SCFVC+ SPoC	0.7649	0.7853	0.7656	Corel 10K dataset	SCFVC+ SPoC	0.7449	0.7592	0.7384
	SpoC +NMF	0.7749	0.7948	0.7895		SpoC +NMF	0.7773	0.7963	0.7738
	SCFVC+ SpoC+ ℓ_0 -NMF	0.7992	0.8156	0.8067		SCFVC+ SpoC+ ℓ_0 -NMF	0.7943	0.8128	0.994
	WDCD + LTpP	0.4287	0.4463	0.4374		WDCD + LTpP	0.4457	0.4618	0.4478
	PZCDM + SPTF	0.4749	0.4835	0.4784		PZCDM + SPTF	0.4827	0.4436	0.4739
DOA-Cluster	SCFVC+ SPoC	0.5938	0.6014	0.5728	DOA-Cluster	SCFVC+ SPoC	0.6321	0.6584	0.6349
	SpoC +NMF	0.6284	0.6369	0.6158		SpoC +NMF	0.6539	0.6673	0.6462
	SCFVC+ SpoC+ ℓ_0 -NMF	0.6495	0.6532	0.6384		SCFVC+ SpoC+ ℓ_0 -NMF	0.6791	0.6815	0.6638
	WDCD + LTpP	0.4572	0.4758	0.4605		WDCD + LTpP	0.4748	0.4909	0.4767
	PZCDM + SPTF	0.5038	0.5129	0.5077		PZCDM + SPTF	0.5119	0.5229	0.50457
Oxford dataset	SCFVC+ SPoC	0.6240	0.6320	0.6030	Corel 10K dataset	SCFVC+ SPoC	0.6632	0.6890	0.6653
	SpoC +NMF	0.6578	0.6673	0.6464		SpoC +NMF	0.6843	0.6968	0.6756
	SCFVC+ SpoC+ ℓ_0 -NMF	0.6791	0.6827	0.6677		SCFVC+ SpoC+ ℓ_0 -NMF	0.7088	0.7109	0.6948

Table 8 continued

Algorithm used	Descriptors used	MAP	P@20	Avg R-precision	Algorithm used	Descriptors used	MAP	P@20	Avg R-precision
QSRF	WDCD + LTP	0.4790	0.4972	0.4869	QSRF	WDCD + LTP	0.4961	0.5109	0.4981
	PZCDM + SPTF	0.5255	0.5340	0.5279		PZCDM + SPTF	0.5330	0.5442	0.5245
	SCFVC+ SPoC	0.6442	0.6520	0.6231		SCFVC+ SPoC	0.6817	0.7079	0.6852
Proposed Algorithm	SpOC +NMF	0.6779	0.6872	0.6663	Proposed algorithm	SpOC +NMF	0.7042	0.7166	0.6955
	SCFVC+ SpOC+ ℓ_0 - NMF	0.6974	0.7029	0.6891		SCFVC+ SpOC+ ℓ_0 - NMF	0.7288	0.7323	0.7141
	WDCD + LTP	0.5287	0.5463	0.5374		WDCD + LTP	0.5457	0.5618	0.5488
	PZCDM + SPTF	0.5749	0.5835	0.5784		PZCDM + SPTF	0.5827	0.5936	0.5739
	SCFVC+ SPoC	0.6938	0.7014	0.6728		SCFVC+ SPoC	0.7321	0.7584	0.7349
	SpOC +NMF	0.7284	0.7669	0.7158		SpOC +NMF	0.7539	0.7673	0.7462
	SCFVC+ SpOC+ ℓ_0 -NMF	0.7485	0.7532	0.7384	SCFVC+ SpOC+ ℓ_0 - NMF	0.7791	0.7815	0.7638	

ing the retrieval performance on all the four datasets, the proposed rank list fusion scheme on average achieved 5% improvement in mAP, 5% improvement in P@20 and 5% improvement in average R-Precision values as compared to baseline approaches. Thus, it can be concluded that the PSO-based rank list fusion scheme works better than the baseline approaches.

7.7 Statistical significance test

In order to assess whether the proposed retrieval method performs better than the baseline models, it is necessary to apply a test of significance. In practice, the test of significance provide information about whether the observed difference in the evaluation scores of various retrieval methods are really meaningful. Based on statistical evidence, the tests of significance determine whether the difference in evaluation scores are not caused by chance or due to inherent noise in the evaluation. A number of different statistical tests are proposed in the literature to determine whether the difference in performance between retrieval methods are significant or not. Among them, Friedman significance test [67] is the most commonly used one and is generally applied to the mean average precision (mAP) of various retrieval models to compare the significance of each of their retrieval results.

The Friedman test is a nonparametric statistical significance test such that it does not make any assumption about the distribution of the measurements and their error. To perform the significance test a null hypothesis is initially framed. In the case of image retrieval, the typical null hypothesis states the equality of different retrieval models. That is, there is no significant difference in the retrieval models selected for evaluation. The Friedman test assumes that there are c retrieval models ($c \geq 2$) to be evaluated and the evaluation scores corresponding to each model are arranged in b rows where b represents the number of datasets. The Friedman test proceeds as follows: initially, different retrieval models are ranked separately for each dataset in such a way that the best performing algorithm gets a rank of 1, the second best algorithm gets a rank of 2, and so on. Then, the total rank of each retrieval model across all the datasets are computed as follows:

$$r_j = \sum_{i=1}^b r_{ij} \quad (24)$$

where r_{ij} is the rank associated with the j -th retrieval model for the i -th dataset.

Finally, the Friedman test statistics F_s is computed as follows:

$$F_s = \frac{12}{b * c * (c + 1)} \sum_{j=1}^c r_j^2 - 3 * b * (c + 1) \quad (25)$$

Fig. 6 Friedman test results for re-ranking algorithms

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	20	3	6.66667	12	0.0074
Error	0	9	0		
Total	20	15			

Test for column effects after row effects are removed

Fig. 7 Friedman test results for rank aggregation algorithms

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	19.125	3	6.375	11.77	0.0082
Error	0.375	9	0.04167		
Total	19.5	15			

Test for column effects after row effects are removed

where r_j^2 is the square of the total of the ranks for j -th retrieval model.

The Friedman test statistics F_s follows a χ^2 distribution with $(c - 1)$ degrees of freedom and having a p -value associated with it. In practice, the p -value is a probability that measures the evidence against the null hypothesis and a lower p -value provides stronger evidence against the null hypothesis. Thus, the null hypothesis can be rejected when the p -value obtained is less than the selected significance level α .

Friedman test results for the mAP values of the proposed and baseline approaches for image re-ranking and rank list fusion while considering only the top performing image descriptors are presented in Figs. 6 and 7. In both the cases, the significance level (α) is set as 0.05, the number of models compared (c) is four and the number of datasets evaluated (b) is also four. From the results shown in Figs. 6 and 7, it is evident that the Friedman test utilizing χ^2 distribution with three degrees of freedom yield 0.0074 and 0.0082 as its respective p -values. In both the cases, the p -values are observed to be lesser than the predefined significance level 0.05. Therefore, the null hypothesis at the significance level $\alpha = 0.05$ can be rejected and it can be concluded that there is remarkable difference between the proposed approaches and the baseline models for the task of image re-ranking and rank aggregation.

7.8 Evaluation of the re-ranking and rank aggregation-based combination strategy

The experimental results summarized in this section illustrates how the proposed strategy for combining image re-ranking results of multiple descriptors using rank aggregation improves the overall effectiveness of the retrieval operation. In this paper, the retrieval results of the distance correlation coefficient-based image re-ranking algorithm for various descriptors are integrated with the PSO-based rank aggregation scheme. This combination strategy is evaluated for different descriptors and datasets. We examined four datasets and for evaluation purpose.

The average MAP values obtained for all the descriptors considered for evaluation in various datasets when the distance correlation coefficient-based image re-ranking algorithm is used in isolation and in combination with PSO-based rank aggregation scheme for the retrieval task is presented in Table 9. As we can observe, the proposed combination strategy yields higher MAP score with remarkable gains for all the descriptors. It should be noted that the proposed combination framework on an average accomplished relative gains of 35.64, 32.64, 35.54 and 34.95% in MAP values on INRIA Holiday [41], Scene 15 [63], Oxford [64] and Corel 10K [65] image collections. The relative gain is estimated by comparing the retrieval score of the proposed rank list fusion scheme and the highest score among the individual descriptors.

Table 9 MAP score obtained with the proposed combination strategy for all the four datasets

Dataset used	Before re-ranking				After re-ranking				After aggregating re-ranking results	Gain (%)
	OB	SCFVC	SPoC	ℓ_0 -NMF	OB	SCFVC	SPoC	ℓ_0 -NMF		
INRIA Holiday dataset	0.5732	0.6081	0.6314	0.6480	0.6873	0.71583	0.7436	0.7719	0.8790	35.64
Scene 15 dataset	0.5544	0.5736	0.6172	0.6430	0.6727	0.7065	0.7383	0.7662	0.8529	32.64
Oxford dataset	0.5271	0.5422	0.5639	0.5930	0.6363	0.6667	0.6683	0.7275	0.8038	35.54
Corel 10K dataset	0.5684	0.5872	0.6074	0.6279	0.6881	0.7129	0.7362	0.7717	0.8474	34.95

Table 10 P@20 values obtained with the proposed combination strategy for all the four datasets

Dataset used	Before re-ranking				After re-ranking				After aggregating re-ranking results	Gain (%)
	OB	SCFVC	SPoC	ℓ_0 -NMF	OB	SCFVC	SPoC	ℓ_0 -NMF		
INRIA Holiday dataset	0.5961	0.6153	0.6541	0.6603	0.7059	0.73149	0.7951	0.8068	0.8810	33.42
Scene 15 dataset	0.5681	0.5919	0.6277	0.6676	0.6828	0.7266	0.7576	0.7948	0.8697	30.27
Oxford dataset	0.5393	0.5514	0.5778	0.6069	0.6548	0.6885	0.7081	0.7447	0.8262	36.13
Corel 10K dataset	0.5747	0.5973	0.6227	0.6373	0.6907	0.7258	0.7574	0.7718	0.8389	31.63

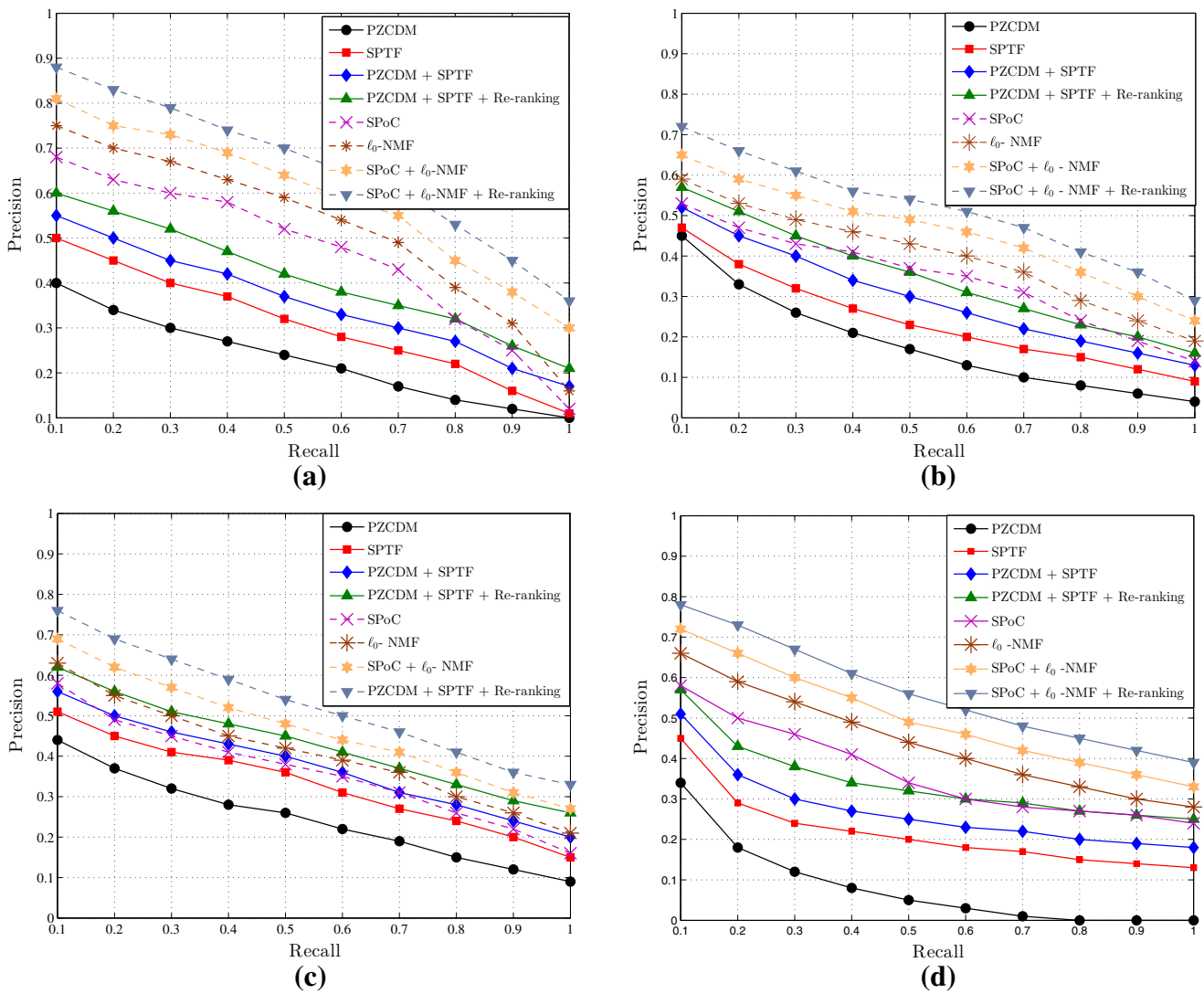


Fig. 8 11-point interpolated average precision values of the proposed combination strategy for image retrieval. **a** INRIA Holiday dataset. **b** Scene-15 dataset. **c** Oxford dataset. **d** Corel 10K dataset

The proposed model is further evaluated on the basis of average Precision@20 (P@20) values and the results are summarized in Table 10. The obtained average P@20 scores indicate the fact that the proposed combination strategy is a promising alternative to image retrieval. In addition, Fig. 8 shows the 11-point interpolated precision values of the proposed framework under varying situations: before and after using the proposed combination strategy for image retrieval. As it can be seen, there is significant gain in precision with the proposed approach in all the retrieval experiments conducted across all the datasets.

8 Conclusion

In this paper, new strategies for image re-ranking and rank aggregation are proposed and are efficiently integrated to

further improve the retrieval performance of existing CBIR systems. The proposed framework unifies a distance correlation coefficient-based image re-ranking algorithm and a PSO-based rank list fusion scheme. This enables the re-ordering of retrieval lists generated by multiple CBIR systems and the aggregation of these fine-tuned results to have an enhanced solution. The proposed framework is evaluated using low-level and high-level image descriptors. A rich set of experiments were conducted and the obtained results demonstrated improved performance in terms of effectiveness and efficiency as compared to the results of individual CBIR systems in isolation. In future, the possibility of combining the proposed framework with certain supervised approaches such as relevance feedback will be investigated.

References

- Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S.: Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Technol.* **8**(5), 644–655 (1998)
- Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: *Proceedings of the Ninth ACM International Conference on Multimedia*, ACM, pp. 107–118 (2001)
- Su, Z., Zhang, H., Li, S., Ma, S.: Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. *IEEE Trans. Image Process.* **12**(8), 924–937 (2003)
- Zhou, X.S., Huang, T.S.: Relevance feedback in image retrieval: a comprehensive review. *Multimed. Syst.* **8**(6), 536–544 (2003)
- Ion, A.L., Stanescu, L., Burdescu, D.: Semantic based image retrieval using relevance feedback. In: *The International Conference on Computer as a Tool EUROCON*, 2007, IEEE, pp. 303–310 (2007)
- Ferecatu, M., Boujemaa, N., Crucianu, M.: Semantic interactive image retrieval combining visual and conceptual content description. *Multimed. Syst.* **13**(5–6), 309–322 (2008)
- Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the web. In: *Proceedings of the 10th international conference on World Wide Web*, ACM, pp. 613–622 (2001)
- Marshall, B., Wilson, D.-M.: Applying aggregation concepts for image search. In: *Tenth IEEE International Symposium on Multimedia (ISM)*, 2008, IEEE, pp. 328–333 (2008)
- Cormack, G.V., Clarke, C.L., Buettcher, S.: Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 758–759 (2009)
- Park, G., Baek, Y., Lee, H.-K.: Re-ranking algorithm using post-retrieval clustering for content-based image retrieval. *Inf. Process. Manag.* **41**(2), 177–194 (2005)
- Hsu, W.H., Kennedy, L.S., Chang, S.-F.: Video search reranking through random walk over document-level context graph. In: *Proceedings of the 15th International Conference on Multimedia*, ACM, pp. 971–980 (2007)
- Kontschieder, P., Donoser, M., Bischof, H.: Beyond pairwise shape similarity analysis. In: *Computer Vision—ACCV 2009*, Springer, pp. 655–666 (2009)
- Tian, X., Yang, L., Wang, J., Wu, X., Hua, X.-S.: Bayesian visual reranking. *IEEE Trans. Multimed.* **13**(4), 639–652 (2011)
- Shen, X., Lin, Z., Brandt, J., Avidan, S., Wu, Y.: Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, IEEE, pp. 3013–3020 (2012)
- Pedronette, D.C.G., Torres, R.S.: Image re-ranking and rank aggregation based on similarity of ranked lists. *Pattern Recognit.* **46**(8), 2350–2360 (2013)
- Porkaew, K., Chakrabarti, K.: Query refinement for multimedia similarity retrieval in mars. In: *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, ACM, pp. 235–238 (1999)
- Deselaers, T., Paredes, R., Vidal, E., Ney, H.: Learning weighted distances for relevance feedback in image retrieval. In: *19th International Conference on Pattern Recognition (ICPR)*, 2008, IEEE, pp. 1–4 (2008)
- Drucker, H., Shahrany, B., Gibbon, D.C.: Relevance feedback using support vector machines. In: *ICML*, pp. 122–129 (2001)
- Renda, M.E., Straccia, U.: Web metasearch: rank versus score based rank aggregation methods. In: *Proceedings of the 2003 ACM Symposium on Applied computing*, ACM, pp. 841–846 (2003)
- Liu, Y., Mei, T., Hua, X.-S., Tang, J., Wu, X., Li, S.: Learning to video search rerank via pseudo preference feedback. In: *IEEE International Conference on Multimedia and Expo*, 2008, IEEE, pp. 297–300 (2008)
- Amir, A., Berg, M., Chang, S.-F., Hsu, W., Iyengar, G., Lin, C.-Y., Naphade, M., Natsev, A., Neti, C., Nock, H. et al.: Ibm research trecvid-2003 video retrieval system. *NIST TRECVID* (2003)
- Liu, Y., Mei, T., Hua, X.-S., Tang, J., Wu, X., Li, S.: Learning to video search rerank via pseudo preference feedback. In: *IEEE International Conference on Multimedia and Expo*, 2008, IEEE, pp. 297–300 (2008)
- Rudinac, S., Larson, M., Hanjalic, A.: Exploiting visual re-ranking to improve pseudo-relevance feedback for spoken-content-based video retrieval. In: *Proceedings of the 10th Workshop on Image Analysis for Multimedia Interactive Service*, IEEE, pp. 17–20 (2009)
- Liu, Y., Mei, T.: Optimizing visual search reranking via pairwise learning. *IEEE Trans. Multimed.* **13**(2), 280–291 (2011)
- Liu, W., Jiang, Y.-G., Luo, J., Chang, S.-F.: Noise resistant graph ranking for improved web image search. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, IEEE, pp. 849–856 (2011)
- Wang, M., Li, H., Tao, D., Lu, K., Wu, X.: Multimodal graph-based reranking for web image search. *IEEE Trans. Image Process.* **21**(11), 4649–4661 (2012)
- Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N.: Query specific rank fusion for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(4), 803–815 (2015)
- Qin, D., Gammeter, S., Bossard, L., Quack, T., Van Gool, L.: Hello neighbor: accurate object retrieval with k-reciprocal nearest neighbors. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, IEEE, pp. 777–784 (2011)
- Jing, Y., Baluja, S.: Visualrank: Applying pagerank to large-scale image search. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(11), 1877–1890 (2008)
- Yang, X., Bai, X., Latecki, L.J., Tu, Z.: Improving shape retrieval by learning graph transduction. In: *Computer Vision—ECCV 2008*, Springer, pp. 788–801 (2008)
- Yang, X., Koknar-Tezel, S., Latecki, L.J.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, IEEE, pp. 357–364 (2009)
- Yang, X., Latecki, L.J.: Affinity learning on a tensor product graph with applications to shape and image retrieval. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, IEEE, pp. 2369–2376 (2011)
- Pedronette, D.C.G., Torres, R.S.: Exploiting clustering approaches for image re-ranking. *J. Vis. Lang. Comput.* **22**(6), 453–466 (2011)
- Pedronette, D.C.G., Penatti, O.A., Torres, R.S.: Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks. *Image Vis. Comput.* **32**(2), 120–130 (2014)
- Aslam, J.A., Montague, M.: Models for metasearch. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 276–284 (2001)
- Lu, T., Boutilier, C.: Learning mallows models with pairwise preferences. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 145–152 (2011)
- Qin, T., Geng, X., Liu, T.-Y.: A new probabilistic model for rank aggregation. In: *Advances in Neural Information Processing Systems*, pp. 1948–1956 (2010)
- Fox, E.A., Shaw, J.A.: Combination of multiple searches, *NIST SPECIAL PUBLICATION SP* pp. 243–243 (1994)
- Jain, A.K., Vailaya, A.: Image retrieval using color and shape. *Pattern Recognit.* **29**(8), 1233–1244 (1996)

40. Depeursinge, A., Müller, H.: Fusion techniques for combining textual and visual information retrieval. In: Image CLEF, Springer, pp. 95–114 (2010)
41. Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. In: Computer Vision—ECCV 2008, Springer, pp. 304–317 (2008)
42. Wang, X.-Y., Zhang, B.-B., Yang, H.-Y.: Content-based image retrieval by integrating color and texture features. *Multimed. Tools Appl.* **68**(3), 545–569 (2014)
43. Székely, G.J., Rizzo, M.L., Bakirov, N.K., et al.: Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35**(6), 2769–2794 (2007)
44. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **6**(2), 182–197 (2002)
45. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
46. Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **11**(4), 341–359 (1997)
47. Karaboga, D., Akay, B.: A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
48. Yang, X.-S., Deb, S.: Engineering optimisation by cuckoo search. *Int. J. Math. Model. Numer. Optim.* **1**(4), 330–343 (2010)
49. Ab Wahab, M.N., Nefti-Meziani, S., Atyabi, A.: A comprehensive review of swarm optimization algorithms. *PloS ONE* **10**(5), e0122827 (2015)
50. Talib, A., Mahmuddin, M., Husni, H., George, L.E.: A weighted dominant color descriptor for content-based image retrieval. *J. Vis. Commun. Image Represent.* **24**(3), 345–360 (2013)
51. Tran, L.V., Lenz, R.: Compact colour descriptors for colour-based image retrieval. *Signal Process.* **85**(2), 233–246 (2005)
52. Murala, S., Maheshwari, R., Balasubramanian, R.: Directional local extrema patterns: a new descriptor for content based image retrieval. *Int. J. Multimed. Inf. Retr.* **1**(3), 191–203 (2012)
53. Murala, S., Maheshwari, R., Balasubramanian, R.: Local tetra patterns: a new feature descriptor for content-based image retrieval. *IEEE Trans. Image Process.* **21**(5), 2874–2886 (2012)
54. Ricard, J., Coeurjolly, D., Baskurt, A.: Generalizations of angular radial transform for 2d and 3d shape retrieval. *Pattern Recognit Lett* **26**(14), 2174–2186 (2005)
55. Zhang, D., Lu, G.: Generic fourier descriptor for shape-based image retrieval. In: Proceedings of the IEEE International Conference on Multimedia and Expo, 2002 ICME'02, vol. 1, IEEE, pp. 425–428 (2002)
56. El-ghazal, A., Basir, O., Belkasim, S.: Invariant curvature-based fourier shape descriptors. *J. Vis. Commun. Image Represent.* **23**(4), 622–633 (2012)
57. Chum, O., Philbin, J., Zisserman, A. et al.: Near duplicate image detection: min-hash and tf-idf weighting. In: *BMVC*, vol. 810, pp. 812–815 (2008)
58. Jégou, H., Douze, M., Schmid, C., Pérez, P.: Aggregating local descriptors into a compact image representation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, IEEE, pp. 3304–3311 (2010)
59. Li, L.-J., Su, H., Lim, Y., Fei-Fei, L.: Object bank: an object-level image representation for high-level visual recognition. *Int. J. Comput. Vis.* **107**(1), 20–39 (2014)
60. Liu, L., Shen, C., Wang, L., van den Hengel, A., Wang, C.: Encoding high dimensional local features by sparse coding based fisher vectors. In: *Advances in Neural Information Processing Systems*, pp. 1143–1151 (2014)
61. Zhao, B., Xing, E.: Sparse output coding for large-scale visual recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3350–3357 (2013)
62. Arun, K.S., Govindan, V.: Optimizing visual dictionaries for effective image retrieval. *Int. J. Multimed. Inf. Retr.* **4**(3), 165–185 (2015)
63. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, vol. 2, IEEE, pp. 2169–2178 (2006)
64. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007 (CVPR'07), IEEE, pp. 1–8 (2007)
65. Liu, G.-H., Yang, J.-Y.: Content-based image retrieval using color difference histogram. *Pattern Recognit.* **46**(1), 188–198 (2013)
66. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: *Robust Statistics: The Approach Based on Influence Functions*, vol. 114. Wiley, Hoboken (2011)
67. Sheskin, D.J.: *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, Boca Raton (2003)