

Combining re-ranking and rank aggregation methods for image retrieval

Daniel Carlos Guimarães Pedronette¹ ·
Ricardo da S. Torres²

Received: 29 November 2014 / Revised: 9 October 2015 / Accepted: 26 October 2015 /
Published online: 12 November 2015
© Springer Science+Business Media New York 2015

Abstract This paper presents novel approaches for combining re-ranking and rank aggregation methods aiming at improving the effectiveness of Content-Based Image Retrieval (CBIR) systems. Given a query image as input, CBIR systems retrieve the most similar images in a collection by taking into account image visual properties. In this scenario, accurately ranking collection images is of great relevance. Aiming at improving the effectiveness of CBIR systems, *re-ranking* and *rank aggregation* algorithms have been proposed. However, different re-ranking and rank aggregation approaches, applied to different image descriptors, may produce different and complementary image rankings. In this paper, we present four novel approaches for combining these rankings aiming at obtaining more effective results. Several experiments were conducted involving shape, color, and texture descriptors. The proposed approaches are also evaluated on multimodal retrieval tasks, considering visual and textual descriptors. Experimental results demonstrate that our approaches can improve significantly the effectiveness of image retrieval systems.

Keywords Content-based image retrieval · Re-ranking · Rank aggregation · Fusion

✉ Daniel Carlos Guimarães Pedronette
daniel@rc.unesp.br

Ricardo da S. Torres
rtorres@ic.unicamp.br

¹ Department of Statistics, Applied Mathematics and Computing, São Paulo State University (UNESP), Rio Claro, Brazil

² Recod Lab - Institute of Computing, University of Campinas (UNICAMP), Campinas, Brazil

1 Introduction

Currently, image and multimedia retrieval consists in a research challenge of great interest in academia and industry. Successive advances in image acquisition and sharing technologies have enabled significant changes in the users' behavior. Users have moved from consumers to active producers of multimedia content. In this scenario, effective systems for analyzing and searching images by content are mandatory. Content-Based Image Retrieval (CBIR) can be broadly seen as any technology that helps to organize digital picture archives by means of their visual content [11]. In general, given a query image, a CBIR system aims at retrieving the most similar images in a collection. Images are ranked in decreasing order of similarity between the feature vectors of the query image and dataset images. The descriptors take into account different visual properties of the image (e.g., shape, color, and texture).

Although very promising and present in various industry retrieval system, the current approaches also have limitations. A common problem consists in the use of visual similarity for judging semantic similarity. This limitation is generally known as “*semantic gap*,” because is related to the mismatch between low-level features and high-level semantic concepts [11]. The semantic gap, in turn, has direct and negative impact on the effectiveness of retrieval systems. Recently, post-processing approaches, as *re-ranking* algorithms [41], have been proposed in order to improve the effectiveness of retrieval results, without the need of user intervention.

The research involving visual features is also very active and several different features [17, 20, 27, 31] have been proposed, considering shape, color, and texture properties. In this scenario, an alternative solution for improving the effectiveness of retrieval systems consists in combining different visual features using *rank aggregation* methods. For a retrieval task, different features produce different rankings. Therefore, it is intuitive that different descriptors may provide different but complementary information about images. In this way, rank aggregation techniques [3, 7, 15, 38] aim at combining different rankings in order to obtain a more accurate one. Although diverse rank aggregation approaches have been proposed for years [16], this research area is still attracting attention of CBIR systems researches [38].

More recently, there has been considerable research on contextual re-ranking methods. In general, CBIR systems consider only pairwise image analysis, that is, compute similarity measures considering only pairs of images, ignoring available information encoded in the relations among several images. On the other hand, the user perception usually considers the query specification and responses in a given *context*. Therefore, the main objective of these methods consists in exploiting the contextual information aiming at improving the distance measures. Significant effectiveness gains have been obtained, considering various algorithms and techniques [21, 22, 32, 34, 36, 38, 41, 43, 49, 62, 63].

In the same way that different visual features produce different and complementary rankings, results of *re-ranking* and *rank aggregation* methods can also be combined to obtain more effective results. Although a lot of efforts have been employed to develop new visual features, re-ranking, and rank aggregation methods, few initiatives aim to combine them. In this paper, we propose four novel approaches for combining re-ranking and rank aggregation methods aiming at improving the effectiveness of CBIR systems. We discuss how to combine (i) re-ranking algorithms; (ii) rank aggregation algorithms, and both (iii) re-ranking and rank aggregation algorithms.

A large experimental evaluation was conducted, considering different datasets and various image descriptors (using shape, color, and texture properties). We evaluated the proposed approaches on multimodal retrieval tasks, considering visual and textual descriptors, and performed comparisons with stated-of-the-art methods. Experimental results demonstrate that our combination approaches can further improve the effectiveness of image retrieval systems. This paper differs from previous work [38, 39] as it presents: (i) a formal definition of the proposed methods; (ii) a new meta-agglomerative approach for combining rank aggregation methods; and (iii) extends the experimental protocol, evaluating the proposed method on visual and multimodal image retrieval tasks.

This paper is organized as follows. Section 2 discusses related work. Section 3 presents a formal definition of the adopted image retrieval model using re-ranking and rank aggregation methods. Section 4 presents our approaches for combining re-ranking and rank aggregation methods. Section 5 briefly presents the re-ranking and the rank aggregation methods considered in our experiments. Section 6 presents the experimental evaluation and, finally, Section 7 discusses the conclusions.

2 Related work

This section discusses related work. Section 2.1 presents re-ranking approaches, while Section 2.2 discusses upon rank aggregation methods.

2.1 Re-ranking

In a general retrieval scenario, given a query image, a CBIR system aims at retrieving the most similar images in a collection ranked in decreasing order of similarity. Traditional CBIR systems compute a predefined distance (or similarity) measure between the query image and each collection image. Next, the images are ordered according to the considered measure. Therefore, these approaches perform only pairwise image analysis, that is, they compute measures considering only pairs of images.

However, the traditional distance measures (as the classical Euclidean distance) often fail to return satisfactory results, mainly due to the well-known semantic gap problem [18]. In addition, the pairwise analysis ignores existing information encoded in the relationships among several images. In recent years, there has been considerable research on improving the distance measures in CBIR systems and several approaches [22, 36, 62, 63] have been proposed to improve the effectiveness of retrieval tasks by replacing pairwise similarities by more global affinity measures [63]. These approaches propose improving the effectiveness of image searches by exploiting the information about the relationships among images, commonly referred to as contextual information, in an *unsupervised* way. These methods require no user intervention, training or labeled data, and often use iterative strategies to process contextual information. Significant advances also have been achieved by image retrieval and re-ranking based on text queries [24, 53].

Promising results have been obtained by individual methods, yielding very significant effectiveness gains. On the other hand, different techniques have been employed for exploiting contextual information, as diffusion process [61, 62], clustering [32], frequency of patterns [55], graph-based algorithms [22, 63], and approaches based on the analysis of ranked lists [41].

Considering the positive results and the diversity of strategies employed, it is expected that these approaches can be combined. However, to the best of our knowledge, there are no studies about strategies for combining these methods.

2.2 Rank aggregation

In many retrieval applications, different algorithms often provide different and complementary information about the datasets being analyzed. This scenario is similar for content-based image retrieval systems, in which combining various visual features may improve ranking and retrieval effectiveness performance. A common approach for improving CBIR systems consists in using *rank aggregation* techniques. In many situations, rank aggregation has been seen as a way for obtaining a *consensus* ranking when multiple ranked lists are provided for a set of objects. More specifically, rank aggregation approaches aim at combining different rankings in order to obtain a single and more accurate one.

Although rank aggregation problem has a long history [13, 29], and received great attention by the computer community in the last few decades [16], it is still being employed in many new applications [13, 29]. The rank aggregation methods use mainly the information of: (i) the score computed for an image; and (ii) the position (or rank) assigned to an image in a ranked list. Traditional methods as CombSum and CombMNZ algorithms [16] consider the sum of the normalized relevance scores computed by different systems. The Borda count method [8] uses rank information in voting procedures and rank scores are summed up directly.

Rank aggregation methods can be also specially useful for multimedia retrieval, when multimedia objects are composed of different media such as text and image, providing different kinds of upon the same data. Commonly, different rank aggregation approaches consider that images highly ranked in many ranked lists are likely to be relevant [10]. Most of the techniques developed in that context fall in three different categories: early fusion, late fusion, and transmedia fusion. Early fusion approaches consider the feature space designed via a joint model, integrating different features into a single large vector. Late fusion techniques, on the other hand, merge the monomedia similarity information by means of aggregation functions. The aggregation functions exploit information from different rankings for computing a more accurate one. In transmedia approaches, one of the modalities is first used to gather relevant objects and then other modality is employed to aggregate their results [7].

As a result of the diversity of rank aggregation approaches, there are studies about the comparison among different rank aggregation methods [45], although there is no consensus about the best method for all situations. Our objective in this paper is to combine different re-ranking and rank aggregation approaches recently proposed for image retrieval scenarios.

3 Image retrieval model

This section presents formal definitions for *image descriptor*, *re-ranking*, and *rank aggregation*. Those concepts are used along the text.

3.1 Image descriptor

Let $\mathcal{C}=\{img_1, img_2, \dots, img_N\}$ be an *image collection*. Let \mathcal{D} be an *image descriptor* which can be defined [47] as a tuple (ϵ, ρ) , where:

$\epsilon: \hat{I} \rightarrow \mathbb{R}^n$ is a function, which extracts a feature vector v_j from an image \hat{I} .
 $\rho: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a distance function that computes the distance between two images according to the distance between their corresponding feature vectors.

In order to obtain the distance between two images img_i and img_j , it is necessary to compute the value of $\rho(\epsilon(img_i), \epsilon(img_j))$. For simplicity and readability purposes we use the notation $\rho(img_i, img_j)$. The distance $\rho(img_i, img_j)$ among all images $img_i, img_j \in \mathcal{C}$ can be computed to obtain an $N \times N$ distance matrix A , such that $A_{ij} = \rho(img_i, img_j)$.

Given a query image img_q , we can compute a ranked list R_q in response to the query, based on the distance matrix A . The ranked list $R_q = \{img_1, img_2, \dots, img_N\}$ can be defined as a permutation of the collection \mathcal{C} which put the collection images in increasing order of distance ρ . We also can take every image $img_i \in \mathcal{C}$ as a query image img_q , for obtaining ranked lists for each image of the collection.

3.2 Re-ranking

Re-Ranking algorithms aims at exploiting contextual information encoded in the distance matrix A and in the respective ranked lists which can be computed based on distances. We can formally define a re-ranking algorithm as a function f_r that takes as input the distance matrix A and for computing a new and more effective distance matrix A_r :

$$A_r = f_r(A). \quad (1)$$

Given the new distance matrix A_r , new ranked lists can be computed. Note that the function f_r consists in exploiting relationships among images encoded in matrix A and the respective ranked lists.

3.3 Rank aggregation

Let \mathcal{C} be an image collection and let $\mathcal{D} = \{D_1, D_2, \dots, D_m\}$ be a set of m image descriptors. The set of descriptors \mathcal{D} can be used for computing a set of distances matrices $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$. As discussed in previous subsection, for each distance matrix $A_i \in \mathcal{A}$, a set of ranked lists can be computed. The objective of rank aggregation methods is to use the set \mathcal{A} as input for computing a new (and more effective) distance matrix A_a :

$$A_a = f_a(\mathcal{A}). \quad (2)$$

Based on the combined distance matrix A_a , a new set of ranked lists can be computed.

3.4 Efficiency aspects

Although the re-ranking and rank aggregation methods are formally defined in terms of full distance matrices, the methods are also direct supported by ranked list analysis. Although the ranked lists contain distance scores from the entire collection, the top positions of ranked lists are expected to contain the most relevant images related to the query image. Therefore, most of the methods can be extended for performing updates and analysis considering only a subset of ranked list (and the respective distance information).

Recent works [33] demonstrated the scalability of re-ranking methods that exploit this approach. Indexing structures are used for avoiding the need for computing all distances for each ranked list. In addition, parallel and heterogeneous computing environments have been exploited for accelerating re-ranking methods. Since rank-based approaches are less

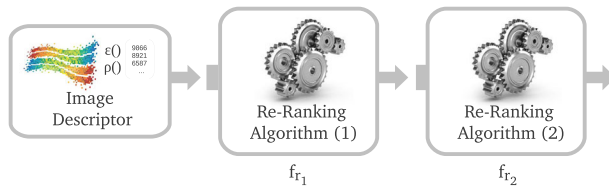


Fig. 1 Cascading re-ranking model involving two re-ranking methods

computationally expensive, recent advances [42] considering parallel architectures achieved real-time performance.

4 Combining re-ranking and rank aggregation methods

This section presents four different approaches for combining re-ranking and rank aggregation methods.

4.1 Cascading re-ranking

In the previous section, we defined a re-ranking algorithm as an implementation of a function $f_r(A)$. Note that in (1), both input and output of the function f_r are distance matrices (the set of ranked lists can be computed based on the distance matrix). In this way, an output matrix obtained from a given function (implemented by a re-ranking algorithm) can be sent to other re-ranking algorithm, aiming at further improving its effectiveness. We call this combination approach as “*cascading re-ranking*,” as it can be applied to a chain of re-ranking algorithms.

More formally, let $\mathcal{F}_r = \{f_{r_1}, f_{r_2}, \dots, f_{r_c}\}$ be a set of functions which represents different re-ranking algorithms. We can use a given distance matrix A for computing a new distance matrix A_r considering a composite function based on the set \mathcal{F}_r as follows:

$$A_r = (f_{r_1} \circ f_{r_2} \circ \dots \circ f_{r_c})(A). \quad (3)$$

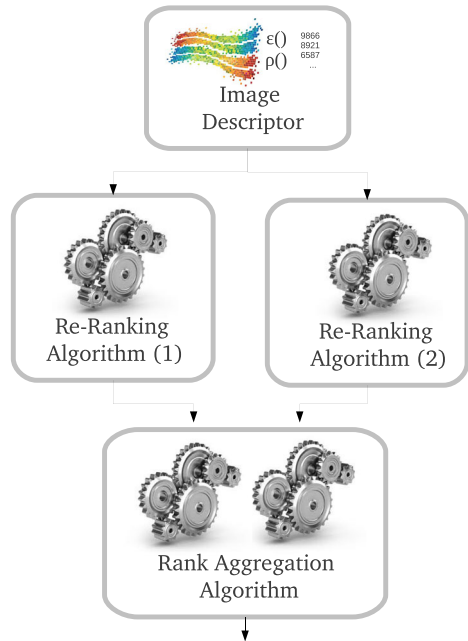
The main motivation of this approach is based on two facts: (i) different re-ranking algorithms exploit contextual information in different ways and can be complementary (one algorithm can improve the quality of ranked lists that others did not); (ii) the second re-ranking algorithm can take advantage of improvements obtained by the first one. Figure 1 illustrates this combination approach, considering two re-ranking algorithms.

4.2 Re-ranking with rank aggregation combination

A single distance matrix A computed by a given image descriptor can be submitted to different re-ranking algorithms, defined by the set of functions $\mathcal{F}_r = \{f_{r_1}, f_{r_2}, \dots, f_{r_c}\}$. In this scenario, a different distance matrix is computed for each re-ranking algorithm, producing a set of matrices $\{A_1, A_2, \dots, A_c\}$.

As discussed in the previous section, the results obtained by the different re-ranking methods are often complementary (one re-ranking can exploit contextual information that

Fig. 2 Rank Aggregation of ranked lists defined by two re-ranking algorithms



others did not). On the other hand, rank aggregation methods aims at combining complementary results. Therefore, we propose to use a rank aggregation method for combining the results of different re-ranking algorithms in order to obtain a single, and more effective distance matrix A_a . Equation (4) defines this combination approach, using a the function f_a representing a rank aggregation method.

$$A_a = f_a(f_{r_1}(A), f_{r_2}(A), \dots, f_{r_c}(A)). \tag{4}$$

Figure 2 illustrates the proposed model, considering the use of two re-ranking methods followed by a rank aggregation step.

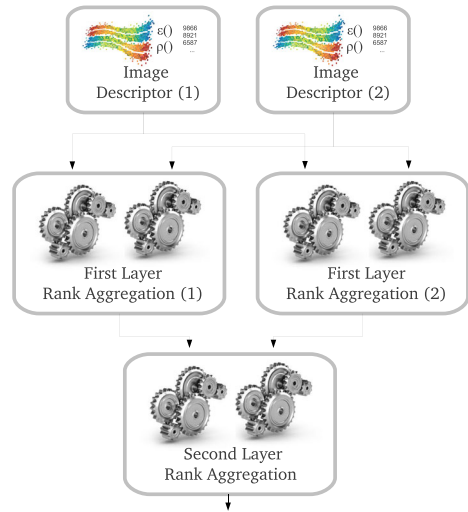
4.3 Agglomerative rank aggregation

The two previous combination approaches consider the strict situation where only one image descriptor is available. In this section, we present a combination approach called “*Agglomerative Rank Aggregation*,” which is able to use several image descriptors as input.

Given a set of distance matrices \mathcal{A} computed by a set of image descriptors, different rank aggregation methods can be employed for combining them. However, each rank aggregation method produces a different and complementary distance matrix. In this way, another rank aggregation method can be employed for combining the results of the first rank aggregation methods. This combination approach uses a hierarchical agglomerative method, in which the rank aggregation methods can be divided in layers.

Considering a combination based on a two-layer approach, the first layer refers to the use of rank aggregation methods that are defined as a set of functions $\mathcal{F}_a = \{f_{a_1}, f_{a_2}, \dots, f_{a_c}\}$. The second layer, in turn, refers to the use of another rank aggregation method, defined by a

Fig. 3 A two-layer rank aggregation scenario. The second-layer rank aggregation method combines the ranked lists defined by two different rank aggregation approaches (first layer)



function f_a . This function combines the matrices computed by the methods used in the first layer. The final distance matrix A_a is computed as follows:

$$A_a = f_a(f_{a_1}(A), f_{a_2}(A), \dots, f_{a_c}(A)). \quad (5)$$

Figure 3 illustrates the proposed approach for a two-layer rank aggregation scenario.

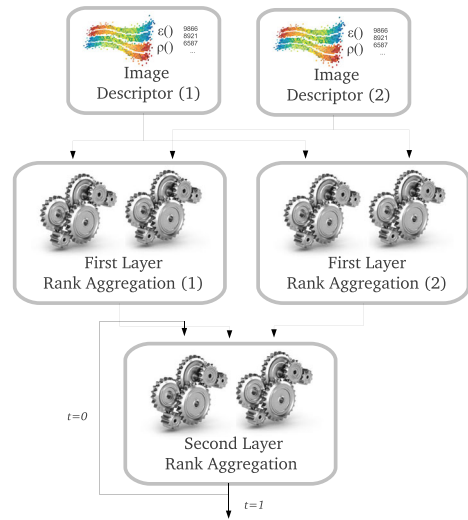
4.4 Meta-agglomerative rank aggregation

As we have previously discussed, the general goal of rank aggregation algorithms is to combine distance matrices computed by different image descriptors aiming at obtaining a new and more effective distance matrix (and ranked lists computed based on distances). In this way, the combined matrix can be considered as a meta-feature and therefore, used as input for the same rank aggregation approach. According to this reasoning, we extend the agglomerative presented in previous section by proposing a *Meta-Agglomerative Rank Aggregation* approach. The first output combined matrix is computed by the agglomerative combination, using the rank aggregation algorithms defined by the first and second layers. In the following, the output matrix is used as an input to the second layer algorithm. Figure 4 illustrates the proposed approach for the meta-agglomerative approach.

5 Re-ranking and rank aggregation methods

This section briefly describes the re-ranking and rank aggregation methods considered in the experimental evaluation of our proposed combination approaches.

Fig. 4 Meta-Agglomerative rank aggregation for a two-layer scenario. The second-layer rank aggregation method combines the ranked lists defined by the first-layer rank aggregation approaches and by the first input ($t=0$) defined by the second-layer approach



5.1 Re-ranking methods

This section presents the two re-ranking methods considered: the Contextual Re-Ranking [34, 40] and the RL-Sim Algorithm [41].

5.1.1 Contextual re-ranking

The Contextual Re-Ranking (CRR) [34, 40] algorithm re-ranks images by taking into account *contextual information* encoded in ranked lists and distance among images. The algorithm uses a *gray scale image* representation of distance matrices computed by CBIR descriptors, referenced as *context image*. The axes of images are defined according to ranked lists and each pixel represent the distance between two images in the dataset, normalized in the interval [0,255] for defining a gray scale.

Given an image, the re-ranking algorithm processes contextual information by constructing *context images* for each one of its *k-nearest neighbors*. The main steps of Contextual Re-Ranking Algorithm are illustrated in Fig. 5.

Each *context image* is analyzed using image processing techniques. Median filters are used to process the *context images*, filtering out wrong similarity information. Subsequently, a binary image is obtained from the gray scale image, where the black pixels represent similarity between the images.

An affinity matrix W is employed for storing the similarity between images, computed using the information of black pixels. More formally, $W[i, j]$ defines the similarity between images img_i and img_j . In this way, the black pixels define increments to the affinity matrix W .

The same procedure is performed for all img_i in the collection. Since all images are processed, the affinity matrix W is used as input for computing a new distance matrix A_{t+1} (where t indicates the current iteration). Based on the new distance matrix A_{t+1} , a new set of ranked lists is computed. These steps are repeated along several iterations. After a

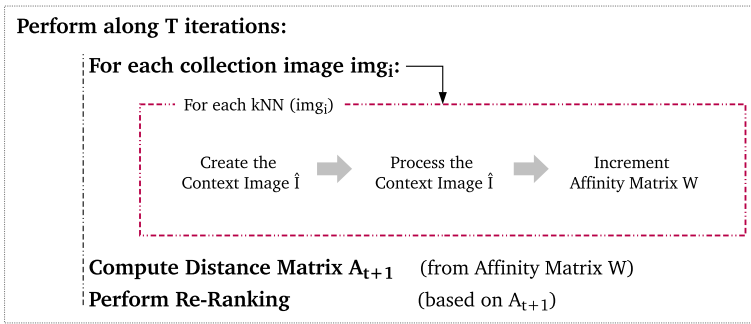


Fig. 5 Main steps of the Contextual Re-Ranking Algorithm

number T of iterations a re-ranking is performed based on final distance matrix A_r in order to obtain the final set of ranked lists.

5.1.2 RL-sim re-ranking algorithm

The main motivation of RL-Sim Re-Ranking [41] algorithm relies on the conjecture that *contextual information encoded in the similarity between ranked lists can provide resources for improving effectiveness of CBIR descriptors*. In general, if two images are similar, their ranked lists should be similar as well [35]. In the following, we present a definition of the re-ranking algorithm.

Let $\psi : \mathcal{R} \times \mathcal{R} \times \mathbb{N} \rightarrow \mathbb{R}$ be a similarity function that defines a similarity score between two ranked lists considering their first K images, such that $\psi(R_x, R_y, K) \geq 0$ for all $R_x, R_y \in \mathcal{R}$. The RL-Sim Re-Ranking [41] takes as input an initial set of ranked lists \mathcal{R} and use the function ψ for computing a more effective distance matrix A_r and therefore a more effective set of ranked lists \mathcal{R}_r . An iterative approach is used: the new (and more effective) set \mathcal{R}_{t+1} , where t indicates the current iteration, is used for the next execution of our re-ranking algorithm and so on. These steps are repeated along several iterations aiming at improving the effectiveness incrementally. After a number T of iterations a re-ranking is performed based on the final distance matrix A_r . Based on final distance matrix, a final set of ranked lists can also be computed.

The objective of function ψ is to compute a more effective distance between two images, considering the contextual information encoded in the first K positions of their ranked lists. Let kNN be a function that extracts a subset with the first positions of ranked list R_x , such that $|kNN(R_x, k)| = k$. The function ψ computes the intersection between the subsets of two ranked lists considering different values of k , such that $k \leq K$. Equation (3) presents the function ψ based on the intersection metric [14].

$$\psi(R_x, R_y, K) = \frac{\sum_{k=1}^K |kNN(R_x, k) \cap kNN(R_y, k)|}{K} \tag{6}$$

A new dissimilarity measure, based on similarity of ranked lists, is computed as follows:

$$\hat{\rho}(img_x, img_y) = \frac{1}{1 + \psi(R_x, R_y, K)} \tag{7}$$

5.1.3 Distance Optimization Algorithm

The Distance Optimization Algorithm (DOA) [36] employs a clustering approach for performing image re-ranking in CBIR tasks. The algorithm is based on the fact that if two images are similar, their distances to other images and therefore their ranked lists should be similar as well. The algorithm creates clusters and uses them for updating distances among images. Based on new distances, more effective clusters can be computed and these steps are repeated in an iterative manner until a convergence criterion is reached.

Clusters are created according to cluster-similar functions, which consider the similarity of their ranked lists or the correlation of their distances. Once clusters are created, distances among images belonging to the same cluster are update (decreased). Redefining distances leads to performing a re-ranking which redefine the clusters. This process is repeated while an unsupervised estimation of the quality of clusters improves.

5.1.4 Mutual kNN graph

The Mutual kNN Graph [22] proposes an unsupervised approach for query-by-example and clustering tasks. The method is evaluated on shape retrieval and goes beyond pairwise similarity analysis by considering the underlying structure of the shape manifold, which is estimated from the similarity scores between all the shapes within a collection.

The manifold structure of the data is exploited by defining a neighborhood for each data point in terms of a modified version of a mutual k -nearest neighbor graph, which yields improved performance on all analyzed datasets [22]. In addition to the modification in the graph construction, possible asymmetries of the input affinity matrix are also exploited.

5.2 Rank aggregation methods

This section briefly describes the rank aggregation approaches considered in our experimental evaluation. We discuss the Contextual Rank Aggregation [37, 40], the use of RL-Sim Algorithm [41] for rank aggregation tasks and the Set Rank Aggregation method.

5.2.1 Contextual rank aggregation

The *Contextual Rank Aggregation (CRA)* [37] algorithm aims at combining the results of different descriptors. The main idea consists in applying the Contextual Re-Ranking [34] algorithm discussed in Section 5.1.1, but using the affinity matrix W for accumulating updates of different descriptors at the first iteration. In this way, different matrices $A_i \in \mathcal{A}$ of different descriptors are being combined.

5.2.2 RL-sim rank aggregation

Let $\mathcal{D} = \{D_1, D_2, \dots, D_m\}$ be a set of CBIR descriptors. We can use the set of descriptors \mathcal{D} for computing a set of distances matrices $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$. The approach used by the RL-Sim Algorithm [41] for combining descriptors works as follows: first, the set \mathcal{A} is combined in a unique matrix A_a . For the matrices combination, a multiplicative approach

is employed. Each position (i, j) of the matrix is computed as follows:

$$A_a[i, j] = (1 + A_1[i, j]) \times (1 + A_2[i, j]) \times \dots (1 + A_m[i, j]) \quad (8)$$

Given a combined matrix A_a , a set of ranked lists \mathcal{R}_a is computed based on this matrix. Then, both the matrix A_a and the set \mathcal{R}_a are used as input for RL-Sim re-ranking algorithm.

5.2.3 Set rank aggregation

In this section, we propose a simple rank aggregation method to be used as a *second layer* rank aggregation in agglomerative approach combination, presented in Section 4.3. We consider the strategy of modeling the ranked lists as *sets* of different sizes. This strategy is also used by the RL-Sim [41] algorithm, presented in Section 5.1.2. We use the same function ψ described in (6) for computing the similarity between ranked lists. We refer to this method as *Set Rank Aggregation (SetRA)* along the paper.

The main idea is to compute the similarity between ranked lists, defined by each distance matrix being combined, and add this similarity scores in order to obtain a new combined score. Let m be the number of matrices being combined and R_{i_x} be the ranked list produced by matrix A_i for image img_x and R_{i_y} for image img_y , the new combined similarity score ψ_c is computed as follows:

$$\psi_c(img_x, img_y, K) = \sum_{i=1}^m \psi(R_{i_x}, R_{i_y}, K) \quad (9)$$

The new combined distance is computed as the inverse of similarity score given by function ψ_c .

6 Experimental evaluation

This section presents a set of conducted experiments for demonstrating the effectiveness of our combination approaches. We analysed our approaches under several aspects and compared our results with other methods from the literature. Section 6.1 describes the datasets and descriptors used for content-based image retrieval tasks. Section 6.2 presents the experimental results concerning the cascading re-ranking methods. Section 6.3 presents results for combination of re-ranking methods. Section 6.4 presents the evaluation of agglomerative rank aggregation approach, while Section 6.5 discusses the evaluation of meta-agglomerative approach. Section 6.6 describes the experimental evaluation of multimodal image retrieval tasks, considering visual and textual descriptors. Section 6.7 discusses performed experiments on the widely used UKBench dataset. Finally, Section 6.8 presents a comparison of our approaches with other re-ranking and rank aggregation methods.

Some sections discuss the gain obtained by the combination approaches. We consider a relative gain formulation aiming at considering the amount of improvement over the initial effectiveness results of each descriptor. Let M_b be the value of the effectiveness measure before the use of the combining algorithms and let M_a be the value after its use, the gain is computed as follows: $Gain = (M_a - M_b)/M_b$.

Table 1 presents a summary of the re-ranking and rank aggregation methods, the sections that they were discussed, and the respective acronyms used along the experimental evaluation. Table 2 shows a summary of datasets used, described in details in next sub-sections. For all datasets, we considered all images as queries in the experimental evaluation.

Table 1 Summary or re-ranking and rank aggregation methods used in the evaluation

Method	Task	Acronym	Section
Contextual Re-Ranking [34]	Re-Ranking	CRR	5.1.1
RL-Sim [41]	Re-Ranking	RL-Sim RR	5.1.2
Distance Optimization Algorithm	Re-Ranking	DOA	5.1.3
Mutual kNN Graph [22]	Re-Ranking	M-kNN Graph	5.1.4
Contextual Rank Aggregation [37]	Rank Aggregation	CRA	5.2.1
RL-Sim [41]	Rank Aggregation	RL-Sim RA	5.2.2
Set Rank Aggregation [39]	Rank Aggregation	SetRA	5.2.3

6.1 Descriptors and datasets

6.1.1 Shape

We use the MPEG-7 collection [25], a well-known shape database, commonly used for re-ranking and post-processing methods evaluation and comparison. This collection is composed of 1400 shapes divided into 70 classes, composed by 20 objects each. In addition to MAP, we consider the bullseye score (*Recall@40*) as evaluation measure, which counts all matching objects within the 40 most similar candidates. The retrieved score is normalized with the highest possible number of hits, which is 20. We consider six shape descriptors using the MPEG-7 collection: Segment Saliences (SS) [48], Beam Angle Statistics (BAS) [1], Inner Distance Shape Context (IDSC) [27], Contour Features Descriptor (CFD) [35], Aspect Shape Context (ASC) [28], and Articulation-Invariant Representation (AIR) [17].

6.1.2 Color

We evaluate our method for three color descriptors: Border/Interior Pixel Classification (BIC) [50], Auto Color Correlograms (ACC) [20], and Global Color Histogram (GCH) [51]. The experiments were conducted on a database used in [58] and composed of images from 7 soccer teams, containing 40 images per class.

6.1.3 Texture

The experiments consider three well-known texture descriptors: Local Binary Patterns (LBP) [31], Color Co-Occurrence Matrix (CCOM) [23], and Local Activity Spectrum

Table 2 Summary of datasets used on the experimental evaluation

Dataset	Type	Size	Relevant per Query
MPEG-7 [25]	Shape	1,400	20
Soccer [58]	Color Scenes	280	40
Brodatz [5]	Texture	1,776	16
UW Dataset [12]	Color Scenes	1,109	variable
UKBench [30]	Objects	10,200	4

Table 3 Cascading re-ranking methods on the MPEG-7 collection (*Recall@40*)

Descriptor	Score	Re-Ranking 1:	Score	Re-Ranking 2:	Cascade Score
CFD [35]	84.43 %	Distance Optimization [35]	92.56 %	Contextual Re-Ranking [34]	93.39 %
CFD [35]	84.43 %	Distance Optimization [35]	92.56 %	RL-Sim Re-Ranking [41]	94.40 %
IDSC [27]	85.40 %	Mutual kNN Graph [22]	93.40 %	Contextual Re-Ranking [34]	93.68 %
IDSC [27]	85.40 %	Mutual kNN Graph [22]	93.40 %	RL-Sim Re-Ranking [41]	94.09 %
CFD [35]	84.43 %	RL-Sim Re-Ranking [41]	94.13 %	Contextual Re-Ranking [34]	94.23 %
CFD [35]	84.43 %	Contextual Re-Ranking [34]	95.71 %	RL-Sim Re-Ranking [41]	95.94 %

(LAS) [52]. We used the Brodatz [5] dataset, a popular dataset for texture descriptors evaluation. The Brodatz dataset are composed of 111 different textures. Each texture is divided into 16 blocks, such that 1776 images are considered.

6.2 Cascading re-ranking

This section discusses the use of our cascading approach for combining re-ranking methods. The main goal of this experiment is to validate that our proposed approach can be used with different re-ranking methods. In this way, we considered the MPEG-7 collection, which is commonly used for evaluation of post-processing methods.

We considered four different re-ranking approaches: the Distance Optimization Algorithm [35], the Mutual kNN Graph [22], the Contextual Re-Ranking [34], and RL-Sim Re-Ranking [41] algorithms. We also considered the Contextual Re-Ranking [34] and RL-Sim Re-Ranking [41] applied to all algorithms.

Table 3 presents the results for *Recall@40* measure. The CFD [35] descriptor was considered due to its high gains obtained for this descriptor. The IDSC [27] descriptor was considered due to its use by the evaluation of the Mutual kNN Graph [22] algorithm. We can observe positive gains over the first re-ranking method for all combinations, ranging from +0.11 % to +1.99 %. The positive gains indicate that, even with contextual information already exploited by the first re-ranking employed, the second re-ranking can further improve the effectiveness when combined by our cascading approach.

6.3 Combining re-ranking methods with rank aggregation

This section presents the effectiveness evaluation of our approach for combining re-ranking with rank aggregation algorithms. We consider the Contextual Re-Ranking (CRR) [34], the RL-Sim Re-Ranking [41], and the Set Rank Aggregation (SetRA) algorithms. We aim also to evaluate the use of the proposed approach for different descriptors and datasets. We considered three datasets and twelve descriptors, including shape, color, and texture descriptors.

Table 4 presents the MAP scores for RL-Sim Re-Ranking [41] and Contextual Re-Ranking [34] algorithms in isolation, and considering our combination approach. The results obtained by an arithmetic mean combination are also showed as a baseline. As we can observe, for almost all descriptors our combination approach presents a higher MAP score than both baselines, with significant gains. Exceptions are the LBP [31] and LAS [52] descriptors, in which the RL-Sim Re-Ranking [41] yields low gains. However, we should

Table 4 Re-Ranking with rank aggregation combination on CBIR Tasks (*MAP*)

Image Descriptor	Dataset	Score	Re-Ranking 1: RL-Sim RR [41]	Re-Ranking 2: CRR [34]	Rank Aggregation	
					Arithmetic Mean	SetRA
SS [48]	MPEG-7	37.67 %	43.06 %	44.79 %	45.40 %	47.33 %
BAS [1]	MPEG-7	71.52 %	74.57 %	76.60 %	77.02 %	78.31 %
IDSC [27]	MPEG-7	81.70 %	86.75 %	87.39 %	87.79 %	88.66 %
CFD [35]	MPEG-7	80.71 %	88.97 %	92.76 %	91.35 %	92.94 %
ASC [28]	MPEG-7	85.28 %	88.81 %	89.82 %	89.87 %	90.62 %
AIR [17]	MPEG-7	89.39 %	93.54 %	94.49 %	96.05 %	97.15 %
GCH [51]	Soccer	32.24 %	33.66 %	33.02 %	32.78 %	33.78 %
ACC [20]	Soccer	37.23 %	43.54 %	39.86 %	41.33 %	46.60 %
BIC [50]	Soccer	39.26 %	43.45 %	43.04 %	44.15 %	47.27 %
LBP [31]	Brodatz	48.40 %	47.77 %	49.06 %	49.37 %	47.93 %
CCOM [23]	Brodatz	57.57 %	62.01 %	63.67 %	63.51 %	64.20 %
LAS [52]	Brodatz	75.15 %	77.81 %	78.48 %	78.86 %	77.89 %
Average		61.34 %	65.32 %	66.08 %	66.45 %	67.72 %

note that, even for those cases, our combination approach presents a MAP score higher than the worst re-ranking methods (in fact, close to the average result between the two methods).

Our approach also presents a higher average score when compared to both re-ranking algorithms. The combination score (67.72 %) represents a gain of +2.48 % for Contextual Re-Ranking [34] algorithm (66.08 %) and +3.67 % to the RL-Sim Re-Ranking [41] algorithm (65.32 %).

We also considered the bullseye score (*Recall@40*) for shape descriptors on the MPEG-7 collection. Table 5 presents the effectiveness results considering the *Recall@40* measure. Similar results were observed for the MAP measure. Our combination approach also presents better average scores (86.69 %) than both re-ranking algorithms. Note that the RL-Sim Re-Ranking [41] algorithm yields better average *Recall@40* scores than Contextual Re-Ranking [34] algorithm. The contrary occurs considering the MAP scores (Contextual

Table 5 Re-Ranking and rank aggregation combination for shape descriptors on the MPEG-7 collection (*Recall@40*)

Shape Descriptor	Score	Re-Ranking 1: RL-Sim RR [41]	Re-Ranking 2: CRR [34]	Rank Aggregation: SetRA
SS [48]	43.99 %	53.15 %	51.38 %	54.69 %
BAS [1]	75.20 %	82.94 %	82.43 %	83.51 %
IDSC [27]	85.40 %	92.18 %	91.84 %	92.16 %
CFD [35]	84.43 %	94.13 %	95.71 %	95.98 %
ASC [28]	88.39 %	94.69 %	93.07 %	93.80 %
AIR [17]	93.67 %	99.90 %	99.80 %	99.99 %
Average	78.51 %	86.17 %	85.71 %	86.69 %

Table 6 Agglomerative rank aggregation combination for CBIR tasks (MAP)

Descriptor	Dataset	Rank Aggregation		Score (MAP)
		First Layer	Second Layer	
CFD [35]	MPEG-7	-	-	80.71 %
ASC [28]	MPEG-7	-	-	85.28 %
CFD [35] + ASC [28]	MPEG-7	Arithmetic Mean	-	89.18 %
CFD [35] + ASC [28]	MPEG-7	Multiplicative Approach	-	95.12 %
CFD [35] + ASC [28]	MPEG-7	RL-Sim RA [41]	-	98.75 %
CFD [35] + ASC [28]	MPEG-7	CRA [37]	-	98.77 %
CFD [35] + ASC [28]	MPEG-7	RL-Sim RA + CRA	SetRA	99.41 %
CFD [35] + ASC [28]	MPEG-7	RL-Sim RA + CRA + Agglomerative RA	SetRA	99.54 %
ACC [20]	Soccer	-	-	37.23 %
BIC [50]	Soccer	-	-	39.26 %
BIC [50] + ACC [20]	Soccer	Arithmetic Mean	-	38.51 %
BIC [50] + ACC [20]	Soccer	Multiplicative Approach	-	38.98 %
BIC [50] + ACC [20]	Soccer	RL-Sim RA [41]	-	44.49 %
BIC [50] + ACC [20]	Soccer	CRA [37]	-	42.14 %
BIC [50] + ACC [20]	Soccer	RL-Sim RA + CRA	SetRA	49.00 %
BIC [50] + ACC [20]	Soccer	RL-Sim RA + CRA + Agglomerative RA	SetRA	50.71 %
CCOM [23]	Brodatz	-	-	57.57 %
LAS [52]	Brodatz	-	-	75.15 %
LAS [52] + CCOM [23]	Brodatz	Arithmetic Mean	-	79.09 %
LAS [52] + CCOM [23]	Brodatz	Multiplicative Approach	-	78.91 %
LAS [52] + CCOM [23]	Brodatz	RL-Sim RA [41]	-	80.26 %
LAS [52] + CCOM [23]	Brodatz	CRA [37]	-	81.63 %
LAS [52] + CCOM [23]	Brodatz	RL-Sim RA + CRA	SetRA	83.70 %
LAS [52] + CCOM [23]	Brodatz	RL-Sim RA + CRA + Agglomerative RA	SetRA	83.71 %

Re-Ranking [34] yields better MAP scores in Table 4). However, our combination approach presents better scores considering both measures.

6.4 Agglomerative rank aggregation

This section presents the evaluation of our proposed Agglomerative Rank Aggregation Approach, introduced in Section 4.3. We select two descriptors with the best effectiveness results for each visual property (shape, color, and texture). For shape descriptors, we do not consider the AIR [17] descriptor because this descriptor presents a very high effectiveness result by itself and our objective is to evaluate the potential of the combination approach.

Table 6 presents the MAP score of our combination approach. For comparison, we also present the MAP scores for descriptors in isolation and other baselines: the arithmetic mean, the multiplicative approach, and an strategy that uses as the first-layer rank aggregation

Table 7 Descriptors scores (MAP) on UW database

Descriptor	Type	Score (MAP)
GCH [51]	Visual - Color	31.75 %
BIC [50]	Visual - Color	43.46 %
JAC [59]	Visual - Color	52.26 %
QCCH [19]	Visual - Texture	17.81 %
LAS [52]	Visual - Texture	20.44 %
HTD [60]	Visual - Texture	22.61 %
DICE [26]	Textual	50.73 %
OKAPI [46]	Textual	51.68 %
BOW [6]	Textual	48.84 %
COS [2]	Textual	41.80 %
JACKARD [26]	Textual	50.29 %
TF-IDF [2]	Textual	49.25 %

method, the RL-Sim Rank Aggregation [41] and Contextual Rank Aggregation (CRA) [37] algorithms. We can observe that significant gains are obtained by the proposed combination approach when compared with the results of the descriptors and with the first-layer rank aggregation method and baselines.

6.5 Meta-Agglomerative Rank Aggregation

This section discusses the effectiveness performance of the proposed Meta-Agglomerative Rank Aggregation Approach. Similarly to the Agglomerative Rank Aggregation evaluation, we select two descriptors with the best effectiveness scores for each visual property (shape, color, and texture). Table 6 presents the MAP score of this combination approach, in which the first layer rank aggregation consists of the use of RL-Sim RA + CRA + Agglomerative RA.

The MAP score for descriptors in isolation and combined with first-layer rank aggregation method are also presented. We can observe that the proposed approach achieves significant gains. Considering the color experiments, for example, the achieved score (50.71 %) presents a gain of +29.16 % over the best description in isolation (39.26 %)

Table 8 Agglomerative rank aggregation for multimodal image retrieval tasks (MAP as score)

Retrieval Task	Descriptors	Agglomerative Rank Aggregation	Borda [64]	Reciprocal [9]
Visual	All visual descriptors	46.17 %	40.29 %	43.29 %
Textual	All textual descriptors	62.54 %	53.07 %	53.14 %
Multimodal	All descriptors	69.73 %	54.89 %	59.34 %
Visual	BIC+JAC	61.58 %	52.54 %	53.00 %
Textual	DICE+OKAPI	62.64 %	54.57 %	54.31 %
Multimodal	BIC+JAC+ DICE+OKAPI	73.54 %	61.91 %	63.67 %

Table 9 Comparison of N-S scores on the UKBench dataset

ACC [20]	VOC [57]	Jégou et al. [21]	Qin et al. [44]	Wang et al. [56]	Zhang et al. [65]	Agglomerative Rank Aggregation
3.36	3.54	3.68	3.67	3.68	3.77	3.80

6.6 Multimodal retrieval

This section presents the evaluation of Agglomerative Rank Aggregation with regard to a multimodal image retrieval task, which considers both visual and textual descriptors. Descriptors, used dataset, and results are discussed in the following.

6.6.1 UW dataset

The UW dataset [12] was created at the University of Washington and consists of a roughly categorized collection of 1,109 images. The images include vacation pictures from various locations. These images are partly annotated using keywords. On the average, for each image the annotation contains six words. The maximum number of words per image is 22 and the minimum is one. There are 18 categories, ranging from 22 images to 255 images per category.

6.6.2 Descriptors

- *Visual Color Descriptors*: we considered three color descriptors on experiments: Border/Interior Pixel Classification (BIC) [50], Global Color Histogram (GCH) [51] (both already used on Section 6.1.2), and the Joint Correlogram (JAC) [59].
- *Visual Texture Descriptors*: for texture we used the Homogeneous Texture Descriptor (HTD) [60], Quantized Compound Change Histogram (QCCH) [19], and Local Activity Spectrum (LAS) [52] (the last also considered in Section 6.1.3).
- *Textual Descriptors*: six well-known textual similarity measures are considered for textual retrieval, like the Cosine similarity measure (COS), Term Frequency - Inverse Term Frequency (TF-IDF), and the Dice coefficient (DICE).

Table 7 presents the scores (MAP) for mentioned descriptors (visual and textual) on the UW dataset. The best descriptor effectiveness score in terms of MAP achieves 52.26 %.

6.6.3 Results

Experiments were conducted using the proposed Agglomerative Rank Aggregation approach, considering visual, textual, and multimodal retrieval tasks. Two scenarios were evaluated: when all descriptors are used; and when only the best descriptors for each modality are used. Two baselines are considered in the experiments: the traditional Borda [64] method and the recently proposed Reciprocal Rank Fusion [9].

Table 8 presents the MAP results for all conducted experiments. It can be observed that, except for the combination of all visual descriptors (46.27 %), all the remaining results overcome the best individual descriptor (52.26 %). The proposed Agglomerative Rank Aggregation approach yields better results than the baselines methods for all combinations.

Table 10 Re-Ranking methods comparison on the MPEG-7 collection (*Recall@40*)

Algorithm	Descriptor	Score	Gain
Shape Descriptors			
CFD [35]	–	84.43 %	–
IDSC [27]	–	85.40 %	–
ASC [28]	–	88.39 %	–
AIR [17]	–	93.67 %	–
Re-Ranking Methods			
Graph Fusion [65]	CFD [35]	88.20 %	+4.47 %
Graph Fusion [65]	ASC [28]	90.21 %	+2.06 %
LP [61]	IDSC [27]	91.00 %	+6.56 %
DOA [35]	CFD [35]	92.56 %	+9.63 %
LCDP [62]	IDSC [27]	93.32 %	+9.27 %
Mutual kNN [22]	IDSC [27]	93.40 %	+9.37 %
RL-Sim RR [41]	CFD [35]	94.13 %	+7.13 %
CRR [34]	CFD [35]	95.71 %	+13.36 %
LCDP [62]	ASC [28]	95.96 %	+8.56 %
SetRA (CRR+RL-Sim RR)	CFD [35]	95.98 %	+13.68 %
CRR [34]	AIR [17]	99.80 %	+6.54 %
RL-Sim RR [41]	AIR [17]	99.90 %	+6.66 %
TPG [63]	AIR [17]	99.99 %	+6.75 %
SetRA (CRR+RL-Sim RR)	AIR [17]	99.99 %	+6.75 %

The best multimodal retrieval result (73.54 %) presents a gain of **+40.72 %** over the best individual descriptor.

6.7 Natural image retrieval

The Agglomerative Rank Aggregation was also evaluated on a public benchmark dataset used for near-duplicate image retrieval tasks. The UKBench [30] dataset (University of Kentucky Recognition Benchmark, also referred to as Nister and Stewenius (N-S) Dataset) has a total of 10,200 images. The dataset is composed of 2,550 objects or scenes, where each object/scene is captured four times from different viewpoints, distances and illumination conditions. For evaluation purposes, the N-S score is used, computed between 1 and 4, corresponding to the number of relevant images among the first four image returned. The highest achievable score is 4, indicating that all similar images are retrieved at top positions. The UKBench is a very challenging dataset for unsupervised approaches, mainly due to the small number of images in each class.

The experiment considered two descriptors: the Auto Color Correlograms (ACC) [20] and a variant of vocabulary tree based retrieval (VOC) [30, 57], which uses SIFT features. For VOC, we considered the rank positions provided by recent approaches¹ as the distances among images. For the UKBench dataset, we used $k = 5$ for the RL-Sim Rank

¹http://research.rutgers.edu/~shaoting/image_search.html (As of October 2015). Images not present in the provided rankings had their distance defined as a constant $n_s = 200$.

Table 11 Rank aggregation comparison on the MPEG-7 collection (*Recall@40*)

Shape Descriptor	Rank Aggregation	Score [%]
DDGM [54]	–	80.03 %
CFD [35]	–	84.43 %
IDSC [27]	–	85.40 %
SC [4]	–	86.80 %
ASC [28]	–	88.39 %
CFD [35]+IDSC [27]	Borda [64]	91.92 %
CFD [35]+ASC [28]	Borda [64]	93.51 %
CFD [35]+IDSC [27]	Reciprocal [9]	94.98 %
CFD [35]+ASC [28]	Graph Fusion [65]	96.16 %
CFD [35]+ASC [28]	Reciprocal [9]	96.25 %
IDSC [27]+DDGM [54]	Co-Transduction [3]	97.31 %
SC [4]+DDGM [54]	Co-Transduction [3]	97.45 %
SC [4]+IDSC [27]	Co-Transduction [3]	97.72 %
CFD [35]+ASC [28]	CRA [37]	99.38 %
CFD [35]+ASC [28]	RL-Sim RA [41]	99.44 %
CFD [35]+ASC [28]	Agglomerative Rank Aggregation	99.50 %
CFD [35]+ASC [28]	Meta-Agglomerative Rank Aggregation	99.73 %

Aggregation [41], Contextual Rank Aggregation (CRA) [37], and Set Rank Aggregation algorithms.

Table 9 presents the results of the Agglomerative Rank Aggregation in comparison with other image retrieval approaches. The proposed approach achieves a high N-S score of **3.80**, comparable with various state-of-the-art results.

6.8 Comparison with other approaches

We also evaluated our combination approaches in comparison with other re-ranking and rank aggregation methods. We consider several re-ranking and post-processing methods, applied to various shape descriptors. Table 10 presents the results. We consider the re-ranking with rank aggregation combination, presented in Section 4.2, using the RL-Sim Re-Ranking [41] and Contextual Re-Ranking [34] with Set Rank Aggregation method. We can observe that our combination approach achieves high effectiveness results, being comparable to the best scores reported in literature.

Our methods are also compared with other rank aggregation approaches on the MPEG-7 collection. Three baselines are considered: the traditional Borda [64] method, usually used as baseline in several works; the recently proposed Reciprocal Rank Fusion [9] method; and the Co-Transduction [3] method, recently proposed for CBIR applications. In the case of Borda and Reciprocal Rank methods, the same descriptors, which are used with our methods, were combined. Our agglomerative and meta-agglomerative approaches, using the RL-Sim Rank Aggregation [41] and Contextual Rank Aggregation [37] methods were considered for comparison. We can observe that our method outperforms various baselines (Table 11).

Regarding efficiency aspects, the complexity of the rank aggregation methods is defined by the sorting of ranked lists, which leads to $O(N \times N \log(N))$. However, in many situations only the beginning of ranked lists may be processed and parallel models can be employed, leading to very efficient implementations as recently proposed [42].

7 Conclusions

In this paper, we have presented a novel approach for combining different re-ranking and rank aggregation methods, considering four novel combination schemes. The main idea of the our work consists in exploiting complementary rankings computed by different methods in order to obtain more effective results.

We conducted a large experimental evaluation including visual (considering shape, color, and texture descriptors) and multimodal retrieval (considering visual and textual descriptors). Experimental results demonstrate that our approaches can further improve the effectiveness of different image retrieval tasks.

In future work, we intend to investigate the use of approaches for automatically combining the results of re-ranking and rank aggregation methods. We also plan to investigate the use of the proposed methods in image retrieval scenarios in which users provide relevance judgements on retrieved results. The idea is to use these relevance feedbacks to guide the combination of re-ranking and rank aggregation approaches. We also intend to conduct a deeper analysis of the gains obtained by the re-ranking and rank aggregation methods, aiming at identifying patterns which could indicate what descriptors are more suitable for these post-processing methods.

Acknowledgments The authors are grateful to São Paulo Research Foundation - FAPESP (grant 2013/08645-0), CNPq (grants 306580/2012-8 and 484254/2012-0), CAPES, AMD, and Microsoft Research for the financial support. Authors are also grateful to Rodrigo T. Calumby for his support in the experiments involving multimodal searches.

References

1. Arica N, Vural FTY (2003) BAS: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recogn Lett* 24(9–10):1627–1639
2. Baeza-Yates RA, Ribeiro-Neto B (1999) *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston
3. Bai X, Wang B, Wang X, Liu W, Tu Z (2010) Co-transduction for shape retrieval. In: *European conference on computer vision (ECCV'2010)*, vol 3, pp 328–341
4. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4):509–522
5. Brodatz P (1966) *Textures: A Photographic Album for Artists and Designers*. Dover
6. Carrillo M, Villatoro-Tello E, López-López A, Eliasmith C, Montes-Y-Gómez M, Villaseñor Pineda L (2009) Representing context information for document retrieval. In: *8th international conference on flexible query answering systems (FQAS'09)*, pp 239–250
7. Clinchant S, Ah-Pine J, Csurka G (2011) Semantic combination of textual and visual information in multimedia retrieval. In: *ACM international conference on multimedia retrieval (ICMR'11)*, pp 44:1–44:8
8. Coppersmith D, Fleischer LK, Rurda A (2010) Ordering by weighted number of wins gives a good ranking for weighted tournaments. *ACM Transactions Algorithms* 6(3):55:1–55:13
9. Cormack GV, Clarke CLA, Buettcher S (2009) Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In: *ACM SIGIR conference on research and development in information retrieval*, pp 758–759

10. Croft WB (2002) Combining approaches to information retrieval. In: Croft WB, Croft WB (eds) *Advances in information retrieval, The information retrieval*, vol 7, pp 1–36. Springer, US
11. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput Surv* 40(2):5:1–5:60. doi:[10.1145/1348246.1348248](https://doi.org/10.1145/1348246.1348248)
12. Deselaers T, Keysers D, Ney H (2008) Features for image retrieval: an experimental comparison. *Inf Retr* 11(2):77–107
13. Fagin R, Kumar R, Mahdian M, Sivakumar D, Vee E (2004) Comparing and aggregating rankings with ties. In: 23th ACM SIGMOD symposium on principles of database systems (PODS'04), pp 47–58
14. Fagin R, Kumar R, Sivakumar D (2003) Comparing top k lists. In: ACM-SIAM symposium on discrete algorithms (SODA'03), pp 28–36
15. Faria FF, Veloso A, Almeida HM, Valle E, da S. Torres R, Gonçalves MA, Meira Jr. WM (2010) Learning to rank for content-based image retrieval. In: *Multimedia information retrieval (MIR'2010)*, pp 285–294. doi:[10.1145/1743384.1743434](https://doi.org/10.1145/1743384.1743434)
16. Fox EA, Shaw JA (1994) Combination of multiple searches. In: *The Second Text REtrieval Conference (TREC-2)*, NIST Special Publication, vol 500–215, pp 243–252. NIST
17. Gopalan R, Turaga P, Chellappa R (2010) Articulation-invariant representation of non-planar shapes. In: 11th European conference on computer vision (ECCV'2010), vol 3, pp 286–299
18. Hoi SC, Liu W, Chang SF (2010) Semi-supervised distance metric learning for collaborative image retrieval and clustering. *ACM Transactions on Multimedia Computing and Communication Applications* 6(3):18:1–18:26
19. Huang CB, Liu Q (2007) An orientation independent texture descriptor for image retrieval. In: *International conference on communications, circuits and systems (ICCCAS 2007)*, pp 772–776
20. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R (1997) Image indexing using color correlograms. In: *IEEE conference on computer vision and pattern recognition (CVPR'97)*, pp 762–768
21. Jegou H, Schmid C, Harzallah H, Verbeek J (2010) Accurate image search using the contextual dissimilarity measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(1):2–11
22. Kotschieder P, Donoser M, Bischof H (2009) Beyond pairwise shape similarity analysis. In: *Asian conference on computer vision*, pp 655–666
23. Kovalev V, Volmer S (1998) Color co-occurrence descriptors for querying-by-example. In: *International conference on multimedia modeling*, p 32
24. Krapac J, Allan M, Verbeek J, Jurie F (2010) Improving web image search results using query-relative classifiers. In: *IEEE Conference on computer vision and pattern recognition (CVPR'2010)*, pp 1094–1101
25. Latecki LJ, Lakmper R, Eckhardt U (2000) Shape descriptors for non-rigid shapes with a single closed contour. In: *IEEE Conference on computer vision and pattern recognition (CVPR'2000)*, pp 424–429
26. Lewis J, Ossowski S, Hicks J, Errami M, Garner HR (2006) Text similarity: an alternative way to search medline. *Bioinformatics* 22(18):2298–2304
27. Ling H, Jacobs DW (2007) Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(2):286–299. doi:[10.1109/TPAMI.2007.41](https://doi.org/10.1109/TPAMI.2007.41)
28. Ling H, Yang X, Latecki LJ (2010) Balancing deformability and discriminability for shape matching. In: *European conference on computer vision (ECCV'2010)*, vol 3, pp 411–424
29. Liu YT, Liu TY, Qin T, Ma ZM, Li H (2007) Supervised rank aggregation. In: *International Conference on World Wide Web (WWW'2007)*, pp 481–490
30. Nistér D, Stewénius H (2006) Scalable recognition with a vocabulary tree. In: *IEEE Conference on computer vision and pattern recognition (CVPR'2006)*, vol 2, pp 2161–2168
31. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7):971–987
32. Park G, Baek Y, Lee HK (2005) Re-ranking algorithm using post-retrieval clustering for content-based image retrieval. *Inf Process Manag* 41(2):177–194
33. Pedronette DCG, Almeida J, da S, Torres R (2014) A scalable re-ranking method for content-based image retrieval. *Inf Sci* 265(1):91–104
34. Pedronette DCG, da S, Torres R (2010) Exploiting contextual information for image re-ranking. In: *Iberoamerican congress on pattern recognition (CIARP'2010)*, pp 541–548
35. Pedronette DCG, da S, Torres R (2010) Shape retrieval using contour features and distance optimization. In: *International joint conference on computer vision, imaging and computer graphics theory and applications (VISAPP'2010)*, vol 1, pp 197–202
36. Pedronette DCG, da S, Torres R (2011) Exploiting clustering approaches for image re-ranking. *J Vis Lang Comput* 22(6):453–466

37. Pedronette DCG, da S, Torres R (2011) Exploiting contextual information for rank aggregation. In: International conference on image processing (ICIP'2011), pp 97–100
38. Pedronette DCG, da S, Torres R (2011) Exploiting contextual spaces for image re-ranking and rank aggregation. In: ACM international conference on multimedia retrieval (ICMR'11), pp 13:1–13:8
39. Pedronette DCG, da S, Torres R (2012) Combining re-ranking and rank aggregation methods. In: Iberoamerican congress on pattern recognition (CIARP'2012), pp 170–178
40. Pedronette DCG, da S, Torres R (2012) Exploiting contextual information for image re-ranking and rank aggregation. *International Journal of Multimedia Information Retrieval* 1(2):115–128
41. Pedronette DCG, da S, Torres R (2013) Image re-ranking and rank aggregation based on similarity of ranked lists. *Pattern Recogn* 46(8):2350–2360
42. Pedronette DCG, da S, Torres R, Borin E, Breternitz M (2013) RI-sim algorithm acceleration on GPUs. In: International symposium on computer architecture and high performance computing (SBAC 2013), pp 176–183
43. Perronnin F, Liu Y, Renders JM (2009) A family of contextual measures of similarity between distributions with application to image retrieval. In: IEEE conference on computer vision and pattern recognition (CVPR'2009), pp 2358–2365
44. Qin D, Gammeter S, Bossard L, Quack T, van Gool L (2011) Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In: IEEE conference on computer vision and pattern recognition (CVPR'2011), pp 777–784
45. Renda ME, Straccia U (2003) Web metasearch: Rank vs. score based rank aggregation methods. In: Proceedings of the 2003 ACM symposium on applied computing, SAC '03, pp 841–846
46. Robertson SE, Walker S, Jones S, Hancock-Beaulieu M, Gatford M (1994) Okapi at trec-3. In: Text REtrieval Conference, pp 109–126
47. da S, Torres R, Falcão AX (2006) Content-Based Image Retrieval: Theory and Applications. *Revista de Informática Teórica e Aplicada* 13(2):161–185
48. da S, Torres R, Falcão AX (2007) Contour Saliency Descriptors for Effective Image Retrieval and Analysis. *Image Vis Comput* 25(1):3–13
49. Schwander O, Nielsen F (2010) Reranking with contextual dissimilarity measures from representational bregmanl k-means. In: International joint conference on computer vision, imaging and computer graphics theory and applications (VISAPP'2010), vol 1, pp 118–122
50. Stehling RO, Nascimento MA, Falcão AX (2002) A compact and efficient image retrieval approach based on border/interior pixel classification. In: ACM conference on information and knowledge management (CIKM'2002), pp 102–109
51. Swain MJ, Ballard DH (1991) Color indexing. *International Journal on Computer Vision* 7(1):11–32
52. Tao B, Dickinson BW (2000) Texture recognition and image retrieval using gradient indexing. *Journal of Visual Communication and Image Representation* 11(3):327–342
53. Thollard F, Qunot G (2013) Content-based re-ranking of text-based image search results. In: European conference on IR research (ECIR'2013), vol 7814, pp 618–629
54. Tu Z, Yuille AL (2004) Shape matching and recognition - using generative models and informative features. In: European conference on computer vision (ECCV'2004), pp 195–209
55. Voravuthikunchai W, Crémilleux B, Jurie F (2014) Image re-ranking based on statistics of frequent patterns. In: International conference on multimedia retrieval, ICMR'14, p 129
56. Wang B, Jiang J, Wang W, Zhou ZH, Tu Z (2012) Unsupervised metric fusion by cross diffusion. In: IEEE conference on computer vision and pattern recognition (CVPR'2012), pp 3013–3020
57. Wang X, Yang M, Cour T, Zhu S, Yu K, Han T (2011) Contextual weighting for vocabulary tree based image retrieval. In: IEEE international conference on computer vision (ICCV'2011), pp 209–216
58. van (2006) de Weijer, J., Schmid, C.: Coloring local feature extraction. In: European conference on computer vision (ECCV'2006), vol Part II, pp 334–348
59. Williams A, Yoon P (2007) Content-based image retrieval using joint correlograms. *Multimedia Tools and Applications* 34(2):239–248
60. Wu P, Manjunath BS, Newsam SD, Shin HD (1999) A texture descriptor for image retrieval and browsing. In: IEEE workshop on Content-Based access of image and video libraries (CBAIVL'99), pp 3–7
61. Yang X, Bai X, Latecki LJ, Tu Z (2008) Improving shape retrieval by learning graph transduction. In: European conference on computer vision (ECCV'2008), vol 4, pp 788–801
62. Yang X, Koknar-Tezel S, Latecki LJ (2009) 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), pp 357–364

63. Yang X, Latecki LJ (2011) 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), pp 2369–2376
64. Young HP (1974) An axiomatization of borda's rule. *J Econ Theory* 9(1):43–52
65. Zhang S, Yang M, Cour T, Yu K, Metaxas DN (2012) Query specific fusion for image retrieval. In: ECCV, pp 660–673



Daniel Carlos Guimarães Pedronette received a BSc in Computer Science (2005) from the State University of São Paulo (Brazil) and the MSc degree in Computer Science (2008) from the University of Campinas (Brazil). He got his doctorate in Computer Science at the same university in 2012. He is currently an assistant professor at the Department of Statistics, Applied Mathematics and Computing, State University of São Paulo (Brazil). His research interests involves content-based image retrieval, re-ranking, rank aggregation, digital libraries, and image analysis.



Ricardo da S. Torres received the B.Sc. degree in computer engineering from the University of Campinas, Brazil, in 2000. He received the Doctorate in computer science at the same university in 2004. He is an Associate Professor at the Institute of Computing, University of Campinas. His research interests include image analysis, content-based image retrieval, databases, digital libraries, and geographic information systems.