Preprint version of paper published in Neurocomputing journal, 2017.
DOI: http://doi.org/10.1016/j.neucom.2017.04.062

Unsupervised Rank Diffusion for Content-Based Image Retrieval

Daniel Carlos Guimarães Pedronette[1], Ricardo da S. Torres[2]

[1] *Depart. of Statistics, Applied Math. and Computing, State University of São Paulo (UNESP), Rio Claro, SP, Brazil*

[2] *RECOD Lab, Institute of Computing, University of Campinas (UNICAMP), Campinas, SP, Brazil*

Abstract

Despite the continuous development of features and mid-level representations, effectively and reliably measuring the similarity among images remains a challenging problem in image retrieval tasks. Once traditional measures consider only pairwise analysis, context-sensitive measures capable of exploiting the intrinsic manifold structure became indispensable for improving the retrieval performance. In this scenario, diffusion processes and rank-based methods are the most representative approaches. This paper proposes a novel hybrid method, named Rank Diffusion, which uses a diffusion process based on ranking information. The proposed method consists in a diffusion-based re-ranking approach, which propagates contextual information through a diffusion process defined in terms of top-ranked objects, reducing the computational complexity of the proposed algorithm. Extensive experiments considering a rigorous experimental protocol were conducted on six public image datasets and several different descriptors. Experimental results and comparison with state-of-the-art methods demonstrate that high effectiveness gains can be obtained, despite the low-complexity of the algorithm proposed.

Keywords: content-based image retrieval, unsupervised learning, rank diffusion

1. Introduction

Mainly due to significant changes in users' behavior, which moved from being mere consumers to active producers, the process of multimedia content generation suffered profound transformations in last decades [1]. In this scenario, the Content-Based Image Retrieval (CBIR) systems represent a relevant solution, supporting searches capable of taking into account the visual properties of digital image content.

The development of CBIR systems has been mainly supported by the proposal of several visual features, based on different approaches. Many global-, local-, and deep-learning-based features have been proposed, considering different visual properties (shape, color, and texture) and mid-level representations. However, despite the continuous development of features, effectively and reliably measuring the similarity among images remains a challenging problem in image retrieval tasks.

More recently, research initiatives have focused on other stages of the retrieval pipeline [2], which are not directly related to low-level feature extraction or pairwise distance computation. Several initiatives based on supervised learning methods have been exploiting labeled data for improving the retrieval accuracy. Relevance feedback approaches [3, 4], for example, obtain supervised information through user interactions in order to learn distance measures capable of encoding user preferences.

However, in several retrieval scenarios, the training data difficult or infeasible to obtain. In this way, various post-processing methods have been proposed in order to improve the effectiveness of image retrieval on an unsupervised way, without the need of any user intervention. Once traditional measures consider only pairwise analysis, context-sensitive measures capable of exploiting the intrinsic manifold structure became imperative for improving retrieval performance. In general, these post-processing methods aim at replacing pairwise distances by more global affinity measures capable of exploiting the dataset manifold [5].

An effective and widely used approach relies on diffusion processes conducted on affinity graphs [6, 7, 8, 5, 9]. Despite the effectiveness gains, the wide use of post-processing methods on large-scale real-world applications also depends on efficiency and scalability aspects [10]. More recently, due to the high computational costs associated with diffusion-based approaches, other methods have emerged. Mainly based on rank analysis, such contextual rank measures [11, 12, 13, 14] can be efficiently computed.

Diffusion processes and rank-based methods are the most representative categories of post-processing methods, sharing common assumptions that distance between two images should be influenced by the relation among their neighbors on the distance manifold [14]. In this paper, a clear connection between both categories of methods is established. A novel hybrid approach named Rank Diffusion is proposed, using a diffusion process based on ranking information. The diffusion-based approach propagates contextual information defined in terms of top-ranked objects, taking into account the manifold structure. The proposed method establishes a relationship between diffusion approaches and contextual rank measures, since it spreads similarity based on ranking information through a diffusion process.

The capacity of the Rank Diffusion method for considering the geometry of the dataset manifold is illustrated in Figures 1 and 2. Figure 1 illustrates the Two-Moon dataset considering the Euclidean distance. One point is selected as a labeled point in each moon and all other data points are assigned to the closest labeled point. Since the Euclidean distance does not consider the geometry structure of the dataset, the extremities of the moons are misclassified. Figure 2 illustrates the same scenario considering the proposed Rank Diffusion approach. As it can be observed, the ideal classification, which respects the geometry of the whole dataset manifold, is produced.

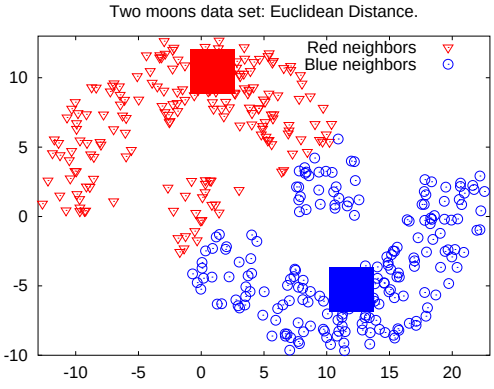

Figure 1: Euclidean distance.

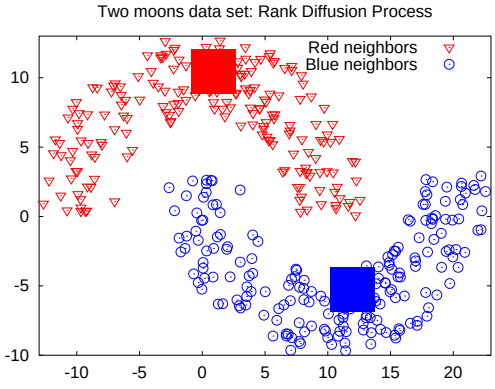

Figure 2: Rank Diffusion.

This work differs from a previously published conference paper [15] in several aspects. A relevant contribution is given in terms of efficiency of the proposed method. Despite the use of a diffusion strategy, a low-complexity re-

ranking algorithm is presented, once only rank information is required. While a traditional diffusion process presents $O(n^3)$ complexity, the Rank Diffusion can be computed in only $O(n)$. Other significant contribution consists in the estimation of an appropriate neighborhood size. Both diffusion [6, 7] and rank-based [11, 14] methods require a k-neighborhood size definition, which are commonly determined empirically. In this paper, a novel approach based on an adaptive neighborhood is proposed, allowing the automatic definition of the neighborhood size, completely independent of ad-hoc parameter settings. In addition, the literature review was deepened and the experimental analysis was significantly extended, including new experiments, descriptors and datasets.

The proposed approach was evaluated through an extensive experimental evaluation, considering six public datasets and several image descriptors, including global (shape, color, and texture), local, and convolution-neural-network-based descriptors. Experiments were conducted on different retrieval tasks, involving shape, color, and texture-based retrieval, object retrieval, and natural image retrieval tasks. The proposed method achieved significant effectiveness gains, yielding very high effectiveness performance in comparison with various state-of-the-art approaches.

The paper is organized as follows: Section 2 and 3 present related work and a formal problem definition, respectively. Section 4 introduces the Rank Diffusion method, while Section 5 discusses how to establish a relationship between diffusion process and rank-based methods. An efficient algorithmic formulation of the Rank Diffusion is discussed in Section 6. Section 7 presents the adaptive neighborhood size selection and Section 8 discusses the rank aggregation approach. Section 9 presents the experimental evaluation and, finally, Section 10 presents the conclusions draws possible future work.

2. Related Work

The continuous growth of image collections made it imperative the development of effective and efficient content-based methods for retrieving the images based on their visual content. However, when dealing with large and diverse collections, i.e., large datasets with diverse content, finding relevant data may become a hard task. The traditional retrieval approaches based only on low-level features, widely used for image and multimedia data applications, sometimes are not able to properly represent data concepts, which

can affect the effectiveness results due to the semantic gap problem [16].

In this scenario, ranking approaches assume a crucial role, capable of incorporating additional information in the retrieval process. In fact, ranking has been established as a relevant task in many diverse domains, including information retrieval, natural language processing, collaborative filtering, and social sciences [17]. One alternative for effectiveness enhancement consists in using learning-to-rank [18] and relevance-feedback [19] strategies, which introduce information about the user perception into ranking models. In learning to rank, groups of objects are given in the training phase, allowing the creation of a ranking model. Next, the ranking model is used to predict the ranked list for a new group of objects [17]. In relevance-feedback approaches, the retrieval results are successively re-computed based on relevance judgments collected from the users.

Although very effective, both learning to rank and relevance feedback strategies are mainly based on supervised learning algorithms, which require user intervention or labeled data. However, in many situations the training data is difficult or infeasible to obtain. In such scenarios, the use of unsupervised learning techniques can represent a promising solution. In a sense, unsupervised learning can be seen as finding patterns in the data what would be considered pure unstructured noise [20]. Besides traditional unsupervised tasks as clustering and dimensionality reduction, the scope of use of unsupervised learning has been expanding from distance learning [21] to feature selection approaches [22].

With the purpose of exploiting the advantages of unsupervised learning methods, many post-processing approaches have been proposed [9, 7, 6, 23, 24, 25, 11, 26] to improve the effectiveness of image retrieval tasks. In general, the objective consists in computing a new distance, which gives rise to more effective ranking results. The main motivation consists in the way of multimedia objects are often modeled and compared. The multimedia content is commonly represented as high-dimensional points in an Euclidean space and the distance between two objects is computed often considering the Euclidean distance. However, once pairwise distance measures define relationships only between pairs of images, the global structure of the dataset and the context wherein the query is computed are ignored. Therefore, how to capture and exploit the intrinsic manifold structure therefore becomes a central problem in the vision and retrieval applications, among others [9]. In this way, such methods take into account the dataset manifold and the global relationships among images for computing new rankings, without the

need of any user intervention.

Some of the most important unsupervised diffusion- and rank-based methods are discussed in the following.

2.1. Diffusion Process Methods

Diffusion is one of the most widely spread processes in science, ranging from motion of dust particles on fluid surfaces to spreading of malaria by migration of mosquitoes [27]. Simply, random walks can be defined as a stochastic process that consists of a sequence of discrete steps of fixed length [28]. Generally speaking, the well-known PageRank algorithm, for example, analyzes the connectivity through a random walk on the link structure of the web.

In the retrieval domain, diffusion-based approaches [29] rely on the definition of a global measure, which describes the relationship between pairs of points in terms of their connectivity. In general, diffusion processes consider as input a pairwise affinity matrix W, which can be interpreted as a graph that encodes the relationships among objects [24]. Let $G = (V, E)$ be a graph, such as the nodes $v_i \in V$ are associated with dataset objects and edges $e_{ij} \in E$ indicate the existence of relationships among them. Edge weights, in turn, are defined by the affinity values w_{ij}. The matrix W is often computed by applying a Gaussian kernel to a distance matrix computed by an image descriptor [9].

Giving the edge weights [24] defined by the matrix W, the diffusion processes spread the affinities through the graph. In general, a walk in the graph occurs more likely through the edges with larger weights. Formally, a probabilistic transition matrix P is defined [9] as a row-wise normalized matrix W. It is also possible to define the probability of being at a specific node after t steps of random walks, considering the initial transition vector and the matrix P^t, where P^t is the power of the matrix P raised to t [24].

Recently, several methods based on diffusion approaches have been proposed. In [7], a metric is learned by collectively propagating the similarity measures through a graph transduction. Graph Transduction (GT) treats the semi-supervised label propagation approach in an unsupervised scenario. In [6], a Locally Constrained Diffusion Process (LCDP) is proposed, considering a local neighborhood for more stability to noise. The Self-Smoothing Operator (SSO) was introduced in [9]. This operator allows computing a new similarity measure directly on the similarity matrix using the self-induced

smoothing kernel. In [24], a generic framework for diffusion processes is proposed in the scope of retrieval applications.

The diffusion-based algorithms have been achieving significant improvements on retrieval performance and are mainly supported by a strong mathematical background. However, such methods are very expensive to compute, specially when the size of datasets becomes larger [12, 13]. The computation of successive steps involving multiplication of the probability matrix are time-consuming operations, which limits the usage of these algorithms in large scale real-world applications.

2.2. Rank-Based Methods

Recently, alternative unsupervised approaches have been proposed for improving the effectiveness of retrieval results by exploiting rank information [11, 12, 13, 14, 26, 30]. In [26], a reciprocal neighborhood analysis is proposed. This method computes different similarity measures for different parts of the ranked lists. The algorithm is mainly based on the observation that the reciprocal neighborhood is a much stronger indicator of similarity than the unidirectional nearest neighborhood relationship. The reciprocal neighborhood is also analyzed by unsupervised manifold learning approaches [12]. The algorithm exploits the top positions of ranked lists and propagates the similarity among neighbors by taking into account the geometry of the dataset manifold.

In [30], a re-ranking method is proposed in which the top-k retrieved images are also used as queries to perform search. By considering those rankings, a new score for each image is collaboratively determined. A ranking consistency method is exploited in re-ranking tasks in [25], where a verification approach refines an initial retrieved ranking list.

An iterative contextual distance measure is proposed in [11]. This method analyzes the similarity among ranked lists. The use of context aims at updating image similarity measures by taking into account information encoded in the top retrieved images. The updated distance can consider different rank correlation measures.

Various contextual rank-based approaches [30, 14, 25, 11, 12, 13] have yielded very significant gains on retrieval effectiveness. Additionally, since the most relevant information of rankings is found at top positions, the rank-based approaches can significantly reduce the computational efforts required

by exploiting indexing structures [10]. Therefore, other important requirements, such as efficiency and scalability [10, 13], are met.

3. Image Ranking Model

Once the proposed method is defined in terms of ranking information, this section presents a formal definition of the image ranking model considered along the paper.

Let $\mathcal{C}$=$\{img_1, img_2, \ldots, img_n\}$ be an image collection, where $n = |\mathcal{C}|$ denotes the size of the collection $\mathcal{C}$. Let $\mathcal{D}$ be an image descriptor which can be defined [31] as a tuple (ϵ, ρ), where:

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

The distance between two images img_i and img_j is computed as $\rho(\epsilon(img_i), \epsilon(img_j))$. The notation $\rho(i, j)$ is used along the paper for readability purposes.

Let img_q be a query image. A ranked list τ_q can be computed in response to img_q based on the distance function ρ. The top positions of ranked lists are expected to contain the most similar images to the query image. Additionally, the full ranked list is expensive to compute, specially when n is high. Therefore, the computed ranked lists can consider only a sub-set of the top-L images.

Let τ_q be a ranked list that contains only the L most similar images to img_q, where $L \ll n$. Formally, let $\mathcal{C}_L$ be a sub-set of the collection $\mathcal{C}$, such that $\mathcal{C}_L \subset \mathcal{C}$ and $|\mathcal{C}_L| = L$. The ranked list τ_q can be defined as a bijection from the set $\mathcal{C}_L$ onto the set $[N] = \{1, 2, \ldots, L\}$.

Every image $img_i \in \mathcal{C}$ can be taken as a query image img_q. As a result, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained, with a ranked list for each image in the collection $\mathcal{C}$. The set $\mathcal{T}$ constitutes a rich source of information about the dataset manifold. While the distance function establishes relationships only between pairs of images, the ranked lists encode the most relevant similarity information about the dataset.

The main purpose of the proposed method consists in exploiting the similarity information encoded in the set $\mathcal{T}$, with the aim of defining a more effective distance function ρ_r. The distance ρ_r, in turn, can be used to compute a new set of ranked lists $\mathcal{T}_r$, which improves the effectiveness of retrieval tasks. Additionally, the fusion problem is also considered, in which various sets of ranked lists $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_d\}$ computed by different descriptors are taken as input, in order to compute a more effective set $\mathcal{T}_r$.

Notice that the input considered by the proposed method is given by ranked lists composed of only a sub-set of with the L most similar images in the collection. In this way, the rank diffusion is completely independent of the initial distance measure ρ and allows that approximate and index-based [32] methods can be used for efficient computation of the ranked lists.

4. Rank Diffusion Process

This section presents the proposed rank diffusion process, presenting each step involved in the method, until the computation of the rank-diffusion distance.

4.1. Rank Similarity Matrix

Many diffusion-based algorithms [24] use the distance information for defining a similarity matrix W. A Gaussian kernel is often considered, such that $w_{ij} = exp(\frac{-\rho^2(i,j)}{2\sigma^2})$, where σ is a parameter to be defined. Therefore, some strategies are required to define a suitable value for the parameter, also considering that the distance distribution may vary among different descriptors.

In this work, a rank similarity matrix W is proposed based only on rank information. The rank modeling of similarity information allows an uniform representation, independent of distance scores. The similarity score w_{ij} varies linearly according to the position of img_j in the ranked list τ_i. Additionally, the score considers only a neighborhood set, which is defined by the size of ranked lists.

Let m denote the size of ranked lists and, therefore the neighborhood considered. Let $\mathcal{N}(i, m)$ be the neighborhood set, which is formally defined as follows:

$$\mathcal{N}(i,m) = \{\mathcal{R} \subseteq \mathcal{C}, |\mathcal{R}| = m \wedge \forall x \in \mathcal{R}, y \in \mathcal{C} - \mathcal{R} : \rho(i,x) \leqslant \rho(i,y)\}. \quad (1)$$

The similarity rank matrix W_m is defined as:

$$w_{m_{ij}} = \begin{cases} m - \tau_i(j) + 1 & if\ img_j \in \mathcal{N}(i,m) \\ 0 & otherwise. \end{cases} \quad (2)$$

The size of the ranked list can assume different values depending on the desired analysis. In the proposed method, the matrix W is defined assuming $m \leq k$, for a local neighborhood analysis, and $m = L$ for a more comprehensive collection sub-set. Notice that, since the matrix W has dimension of $n \times n$, both k and L values are much smaller than n, i.e, $k, L \ll n$. Therefore, the matrix W is very sparse. This property is exploited for defining an efficient algorithm (discussed in Section 6), which computes operations that are equivalent to a matrix multiplication considering W.

4.2. Reciprocal Rank Normalization

While most of similarity pairwise measures are symmetric, the same does not occur for rank measures. In other words, an image img_i well ranked for a query img_j does not imply that img_j is well ranked for query img_i. However, the benefits of improving the symmetry of the k-neighborhood relationship are remarkable in image retrieval applications [23]. Thus, a simple rank normalization procedure is conducted before the rank diffusion process. The reciprocal references among all ranked lists at top-L positions are considered, i.e., $m = L$. For this, the similarity matrix W_L is used and its asymmetry is exploited. The value of w_{ij} is defined considering the position of img_j in the ranked list τ_i, while w_{ji} considers the position of img_i in τ_j. Therefore, a normalized rank similarity matrix $\bar{R}_L$ can be defined as the sum of the original matrix W with its transposed:

$$\bar{R} = W_L + W_L^T. \quad (3)$$

Based on the matrix $\bar{R}$, a rank normalized distance $\bar{\rho}$ is defined as:

$$\bar{\rho}(i,j) = \frac{1}{1 + \bar{r}_{ij}}, \quad (4)$$

where $\bar{r}_{ij} \in \bar{R}$.

In the following, all the ranked lists are updated according to the normalized distance, using a stable sorting algorithm. In this way, all similarity scores defined as 0 in the matrix $\bar{R}$ have their distance changed to 1. In these cases, after the execution of the stable sorting algorithm, the previous order of ranked lists are maintained.

This update gives rise to a new set of ranked lists $\bar{\mathcal{T}}$, used as input for the next steps of the proposed algorithm.

4.3. Iterative Rank Diffusion

The proposed rank diffusion approach is defined by an iterative update of similarity information encoded into a matrix P. The update at each iteration is computed according to a rank similarity matrix W of increasing neighborhood size. The central idea consists in spreading the similarity information through P considering initially a small neighborhood, which is gradually expanded over iterations. Therefore, the number of iterations is defined proportionally to the neighborhood size.

Formally, let $^{(t)}$ denote the current iteration and let s be a constant value, which defines the initial neighborhood size. The rank similarity matrix at a given iteration t is defined as:

$$W^{(t)} = W_{s+t}. \tag{5}$$

where the size of ranked lists $m = (s + t)$. The value of $s = 2$ can be used as a default starting value, or s can be manually defined. The initialization of matrix P is defined considering $t = 0$, and therefore:

$$P^{(0)} = W_s. \tag{6}$$

Given the asymmetry rank relationships, a normalization similarity value is computed proportionally to the accumulated rank similarity of each image. The normalization is defined for matrices W and P, respectively as:

$$\bar{W_{ij}}^{(t)} = \frac{W_{ij}^{(t)}}{\sum_{c=1}^{n} W_{jc}^{(t)}}, \tag{7}$$

and

$$\bar{P}_{ij}^{(t)} = \frac{P_{ij}^{(t)}}{\sum_{c=1}^{n} P_{jc}^{(t)}}. \tag{8}$$

The iterative diffusion step is defined in terms of the multiplication of the normalized matrices $\bar{P}$ and $\bar{W}$. At each iteration, a larger neighborhood is considered in $\bar{W}$ and disseminated along P:

$$P^{(t+i)} = \bar{P}^{(t)} \bar{W^{(t)}}^{T}, \tag{9}$$

where i indicates the increment (its default value is 1). The process is iteratively executed while $t \leq (k-s)$, where k is a parameter that defines the size of the neighborhood considered. The definition of parameter k is discussed in Section 7.

4.4. Reciprocal Rank Diffusion

The diffusion step defined in Equation 9 considers the transposed matrix $\bar{W^{(t)}}^{T}$. In this way, the similarity of a given image img_i to other images is updated according to the ranked list τ_i and is encoded in the similarity matrix.

However, the reciprocal similarity information should also be considered. With this purpose, after the iterative rank diffusion, a self-diffusion step is defined as:

$$P_r = \bar{P}^{(k-s)} \bar{P}^{(k-s)}, \tag{10}$$

where $\bar{P}^{(k-s)}$ represents the last matrix computed by the iterative diffusion after normalization.

4.5. Rank Diffused Distance

Finally, a new rank diffused distance ρ_d is computed inversely proportional to the reciprocal similarity matrix P_r:

$$\rho_r(i,j) = \frac{1}{1 + P_{r_{ij}}} \tag{11}$$

Based on the distance ρ_r, the new and more effective set of ranked lists $\mathcal{T}_r$ is computed using a stable sorting algorithm. As for the rank normalization step, images, which present similarity values equal to 0, maintain the previous order in the ranked lists.

5. Diffusion Process and Contextual Rank Measures

In this section, we briefly discuss the relationship between diffusion process and contextual rank measures, establishing a connection between them. As previously discussed, the matrix W can be interpreted as the edge weights of a graph, which is constrained to k-nearest neighbors in several approaches [24, 6]. The diffusion processes often spread the affinities through the graph computing powers of W (or of a probability transition matrix P computed based on W).

On the other hand, contextual rank approaches compute rank correlation measures exploiting overlap or reciprocal references at top-k positions of ranked lists. In fact, both approaches, although apparently distinct, have much in common.

Consider the example discussed in the following. Let W_k be a rank similarity matrix, as defined in Section 4.1. In a simplified manner, the main step of a diffusion process can be considered as the computation of powers of W_k. However, given the asymmetry of ranked lists, let's consider the multiplication by its transposed for computing a matrix C:

$$C = W_k W_k^T. \tag{12}$$

The matrix C, which represents the basis operations of a diffusion process, can also be interpreted as a rank similarity matrix, where c_{ij} is defined as:

$$c_{ij} = \sum_{r \in \mathcal{C}} w_{k_{ir}} \times w_{k_{jr}}. \tag{13}$$

The value of c_{ij} is different from 0 only if there are overlaps (represented by img_r) between ranked lists τ_i and τ_j at top-k images. For those cases, a larger weight is defined for top positions of ranked lists. This approach is very close to the intersection metric [33] used by contextual rank measures [11].

6. Efficient Diffusion Algorithm

As previously discussed, the diffusion processes based on similarity scores are computationally expensive, requiring the computation of successive matrix multiplication operations. As defined in Section 4, the proposed rank diffusion approach also employs matrix multiplication operations.

However, although defined in terms of matrix operations, the proposed method is entirely based on ranking information. Since the most relevant

information of ranked lists is encoded in top positions, efficient algorithmic solutions can be exploited for computing the rank diffusion process. While the multiplication of similarity scores presents $O(n^3)$ complexity, the rank diffusion can be computed in only $O(n)$. To accomplish this objective, the algorithms used in each step of the proposed method are constrained at constant top positions of ranked lists, defined by constants k and L.

In the following, we present algorithms for efficiently computing the main steps of the proposed method, discussing how the ranking information can be used for reducing the computational efforts.

6.1. Efficient Reciprocal Rank Normalization

The first discussed step consists in the Reciprocal Rank Normalization (Section 4.2), which is defined in terms of sums of matrix W_L, whose complexity is $O(n^2)$. Algorithms 1 and 2 outline an efficient solution for filling the matrix W_L and for computing the Reciprocal Rank Normalization, respectively. Both algorithms consider each ranked list only until the L or $2 \times L$ positions (for the reciprocal analysis) and present a complexity of $O(n)$. The same reasoning is valid for the sorting step.

Algorithm 1 Rank Similarity Matrix

Require: Set of ranked lists $\mathcal{T}$, parameter L
Ensure: Matrix W_L

```
for all img_i ∈ C do
    for all img_j ∈ N(i, L) do
        W_L_ij ← m − τ_i(j)
    end for
end for
```

Notice that the normalized distance ρ computed by the sum of matrices is not strictly equal to the algorithmic value, once matrix elements after top-L positions can be lost. However, in practice, we can observe that such information is not relevant, as experimentally demonstrated in Section 9.2.

The efficient algorithmic solutions of other steps of the method follow the same principle of bounding the processing to the top ranking positions. The sparsity of matrix W (in which less than k positions for each image are non-zero values) is also exploited for drastically reducing the computational complexity.

Algorithm 2 Reciprocal Rank Normalization

Require: Set of ranked lists $\mathcal{T}$, matrix W_L, parameter L
Ensure: Reciprocal normalized set of ranked lists $\bar{\mathcal{T}}$

for all $img_i \in \mathcal{C}$ **do**
  for all $img_j \in \mathcal{N}(i, 2 \times L)$ **do**
    $\bar{R}_{ij} \leftarrow W_{L_{ij}} + W_{L_{ji}}$
    $\bar{\rho}(i, j) \leftarrow 1/(1 + \bar{r}_{ij})$
  end for
end for
$\bar{\mathcal{T}} \leftarrow stableSorting(\mathcal{T}, \bar{\rho})$

Normalization steps defined in Equations 7 and 8, for example, originally use sums involving the entire matrix. However, an efficient algorithm can be employed considering only top-k positions for matrix W normalization and top-L positions for matrix P normalization. Algorithm 3 outlines the proposed W normalization algorithm, restricting the loops of lines 5 and 10 to the top-k positions. The same approach is valid to the matrix P considering the top-L positions.

Algorithm 3 Matrix Normalization

Require: Matrix $W^{(t)}$
Ensure: Normalized Matrix $\bar{W}^{(t)}$

1: **for all** $img_j \in \mathcal{C}$ **do**
2: $s_j \leftarrow 0$
3: **end for**
4: **for all** $img_i \in \mathcal{C}$ **do**
5: **for all** $img_j \in \mathcal{N}(i, k)$ **do**
6: $s_j = s_j + W_{ij}^{(t)}$
7: **end for**
8: **end for**
9: **for all** $img_i \in \mathcal{C}$ **do**
10: **for all** $img_j \in \mathcal{N}(i, k)$ **do**
11: $\bar{W}_{ij}^{(t)} = W_{ij}^{(t)}/s_j$
12: **end for**
13: **end for**

6.2. Efficient Rank Diffusion

The rank diffusion defined by Equation 9, which represents the most expensive step of the method, can also be efficiently computed. Basically, the central idea consists in identifying the non-zero and relevant elements of the matrices W and P based on information of ranked lists. Algorithm 4 presents the proposed approach with a time complexity of $O(n)$.

Algorithm 4 Rank Diffusion Step

Require: Matrices $\bar{W}^{(t)}$ and $\bar{P}^{(t)}$
Ensure: Matrix $P^{(t+1)}$

```
1: for all img_i ∈ C do
2:    for all img_j ∈ N(i, L) do
3:       p_ij^(t+1) ← 0
4:       for all img_l ∈ N(j, k) do
5:          p_ij^(t+1) ← p_ij^(t+1) + (p̄_il^(t) × w̄_jl^(t))
6:       end for
7:    end for
8: end for
```

First, it should be emphasized that the resulting matrix is not fully computed. Instead, only the similarity of the top-L positions are considered (loop of line 2). Additionally, the computational complexity of each element is drastically reduced. More precisely, for computing an element $p_{ij}^{(t+i)}$ using an originally matrix multiplication approach, it is necessary to computed the inner product between two vectors of n elements. Instead, the proposed algorithm computes the product considering only the top-k positions (loop of line 4).

The Reciprocal Rank Diffusion defined in Equation 10 also admits an efficient solution. In fact, the reasoning is very similar to Algorithm 4. However, instead of using the top-k elements of the matrix W, a list of non-zero elements computed in the last iteration of Algorithm 4 is considered.

7. Adaptive Neighborhood Size

Several approaches including both rank-based and diffusion-based methods use a parameter k for determining the size of local neighborhood size. In general, the value of parameter k is empirically or experimentally determined. Although this practice does not invalidate the utility of unsupervised

methods, it represents a limitation. Once unsupervised scenarios imply independence of human intervention, fine tuning actions should be avoided.

In this paper, an adaptive-k approach is proposed for automatically identifying the neighborhood size, turning our method completely independent of parameter settings. More specifically, the proposed method considers four parameters: s, i, k, and L. The parameters s, i and k are directly related to the local neighborhood, which is considered for the rank diffusion: s defines the position of starting neighbor, i represents the increment at each iteration, and k the maximum neighborhood size. Both parameters s and i have natural small default values ($s = 2$ and $i = 1$). On the other hand, the value of L represents a trade-off between effectiveness and efficiency (a detailed discussion is presented in experimental section). Therefore, the parameter k is the only one that needs to be defined a priori.

The neighborhood size is estimated by analyzing the reciprocal references between each image and its neighbors, considering different magnitudes of k. The main motivation consists in determining a value of k where the density of reciprocal references achieves a stable state.

The similarity matrix is used for this analysis, since its squared form W_k^2 is directly associated with reciprocal references. The reciprocal density score $d(q, k)$ of a ranked list τ_q for a neighborhood size k is formally defined as:

$$d(q, k) = \sum_{i \in \mathcal{N}(q,k)} w_{k_{qi}}^2, \tag{14}$$

Considering all ranked lists, an accumulated reciprocal density score $a(k)$ is defined as:

$$a(k) = \sum_{i \in \mathcal{C}} d(i, k). \tag{15}$$

The reciprocal density score is computed considering different iterations, with increasing neighborhood sizes of k. For each size, the variation of the accumulated density score is analyzed, until a stable state is reached. This state is determined in terms of the derivatives of the accumulated reciprocal density score. Let $a^{(k)}$ denote the accumulated authority score at an iteration with neighborhood size k, the first and second derivatives $f_d^{(k)}$ and $s_d^{(k)}$ are defined respectively as:

$$f_d^{(k)} = a^{(k)} - a^{(k-1)}, \tag{16}$$

$$s_d^{(k)} = f_d^{(k)} - f_d^{(k-1)}. \tag{17}$$

The value of k is increased until $s_d^{(k)} < 0$, when the stable state is identified and the loop is interrupted. Section 9.3 presents an experimental analysis of this approach.

8. Rank Aggregation

Different features often encode distinct and complementary visual information extracted from images. Therefore, if a feature produces effective rankings by itself and is complementary (heterogeneous) to other features, then it is expected that a higher search accuracy can be achieved by combining them [34].

In this work, a rank aggregation approach is presented for combining different rankings using the proposed rank diffusion process. The rank diffusion is performed in two stages: first, for each descriptor in isolation and in the following, considering a fused set of ranked lists.

Let $\mathcal{D} = \{D_1, D_2, \ldots, D_d\}$ be a set of different image descriptors and let $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_d\}$ be their respective set of ranked lists. The rank diffusion process is computed for each set $\mathcal{T}_i$, in order to compute a matrix P_r (Equation 10). In the following, a fused matrix P_f is defined as:

$$P_f = \sum_{j \in \mathcal{D}} P_{r_j}. \tag{18}$$

Based on P_f, a new distance ρ_f is computed:

$$\rho_f(i,j) = \frac{1}{1 + P_{f_{ij}}}. \tag{19}$$

A fused set of ranked lists $\mathcal{T}_f$ is computed using the distance ρ_f. Finally, we aim at exploiting the contextual information of the fused set of ranked lists $\mathcal{T}_f$. Once the set $\mathcal{T}_f$ presents the same structure of a set obtained for a single descriptor, it is submitted to the regular rank diffusion process, giving rising to a final set $\mathcal{T}_r$.

9. Experimental Evaluation

The proposed method was evaluated through a rigorous experimental evaluation, considering various datasets, several image descriptors and different image retrieval tasks.

9.1. Datasets and Descriptors

Six public well-known datasets with diverse types of images and diverse characteristics are considered in the experimental evaluation. Several image descriptors are used, including local, global (shape, color, and texture properties), and convolutional neural network-based features. Table 1 presents a summary of datasets and and Table 2 describes the features used for each dataset. The Mean Average Precision (MAP) was used as effectiveness measure for most of experiments. For each dataset, all images are considered as query images, except for Holidays [35] dataset.

Table 1: Datasets used in the experimental evaluation.

Dataset	**Size**	**Type**	**General Description**	**Effectiv. Measure**
MPEG-7 [36]	1,400	Shape	A well known dataset composed of 1400 shapes divided in 70 classes. Commonly used for evaluation of post-processing methods.	MAP, Recall@40
Soccer [37]	280	Color Scenes	Dataset composed of images from 7 soccer teams, containing 40 images per class.	MAP
Brodatz [38]	1,776	Texture	A popular dataset for texture descriptors evaluation composed of 111 different textures divided into 16 blocks.	MAP
ETH-80 [39]	3,280	Objects	Dataset equally divided into 8 classes, with images containing one single object.	MAP
Holidays [35]	1,491	Scenes	Commonly used as image retrieval benchmark, the dataset is composed of 1,491 personal holiday pictures with 500 queries.	MAP
UKBench [40]	10,200	Objects/ Scenes	Popular benchmark, composed of 2,550 objects or scenes. Each object/scene is captured 4 times from different viewpoints, distances, and illumination conditions.	N-S Score

Table 2: Image descriptors considered for each dataset.

Dataset	Image Features	Type
Soccer [37]	Global Color Histogram (GCH) [41], Auto Color Correlograms (ACC) [42], Border/Interior Pixel Classification (BIC) [43]	Color
MPEG-7 [36]	Segment Saliences (SS) [44], Beam Angle Statistics (BAS) [45], Inner Distance Shape Context (IDSC) [46], Contour Features Descriptor (CFD) [47], Aspect Shape Context (ASC) [48], Articulation-Invariant Representation (AIR) [49]	Shape
Brodatz [38]	Local Binary Patterns (LBP) [50], Color Co-Occurrence Matrix (CCOM) [51], Local Activity Spectrum (LAS) [52]	Texture
ETH-80 [39]	ACC [42], BIC [43], GCH [41], and Color Structure Descriptor (CSD) [53]	Color
Holidays [35]	Joint Composite Descriptor (JCD) [54], Scalable Color Descriptor (SCD) [55] Color and Edge Directivity Descriptor Spatial Pyramid (CEED-Spy) [56, 57], ACC [42], Convolutional Neural Network by Caffe [58] (CNN-Caffe), Convolutional Neural Network by OverFeat [59] (CNN-OverFeat)	Color, Texture, BoVW, CNN
UKBench [40]	CEED-Spy [56, 57], Fuzzy Color and Texture Histogram Spatial Pyramid (FCTH-SPy) [60, 57], SCD [55], ACC Spatial Pyramid (ACC-SPy) [42, 57], CNN-Caffe [58] ACC [42], Vocabulary Tree (VOC) [61]	Color, Texture, BoVW, CNN

9.2. Impact of Parameters

This section analyzes the impact parameters on the effectiveness of the proposed method. Experiments were conducted on the MPEG-7 [36] dataset considering the CFD [47] shape descriptor and the MAP as effectiveness measure. The parameters related the local neighborhood are analyzed in the first experiment. The position of starting neighbor was fixed as s=2 while the increment i and the maximum neighborhood size k are varied in the intervals [1,5] and [5,30], respectively. For each combination, we evaluated the obtained MAP. Figure 3 illustrates the results. A large red region can be observed, demonstrating the robustness of the propose method to different parameter settings. For most of experiments, the manual parameter setting was defined as s=5, i=5, and k=20, except for UKBench and Holidays datasets, which used s=2, i=2, k=6 and s=2, i=1, k=3, respectively, due to the small number of images per class. Notice that such parameters apply to a manual setting process, once the adaptive neighborhood approach does not require any parameter.

The second experiment analyzes the trade-off between efficiency and effectiveness given by the constant L. The constant L defines the size of ranked

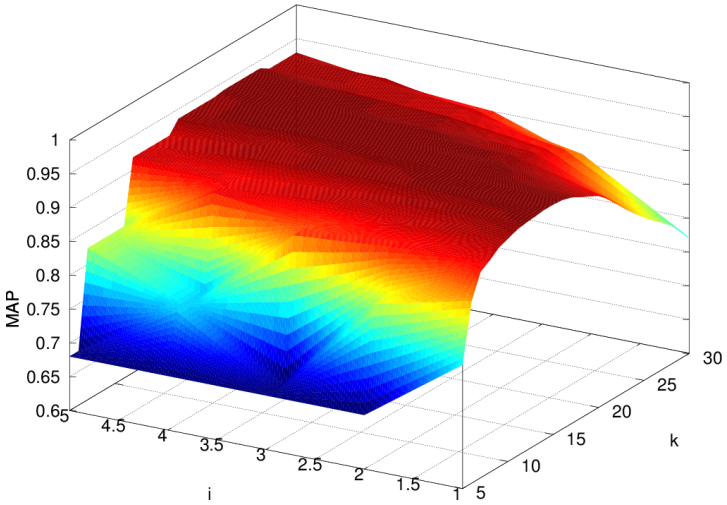

Figure 3: Impact of parameters k and i.

list used as input. The higher the value of L, the higher the effectiveness gains, but the higher the computational costs. Figure 4 shows the impact of L on MAP scores for three shape descriptors on the MPEG-7 [36] dataset. As it can be observed, the most expressive effectiveness gains are obtained for low values of L, which is a very positive property of the algorithm when dealing with large collections.

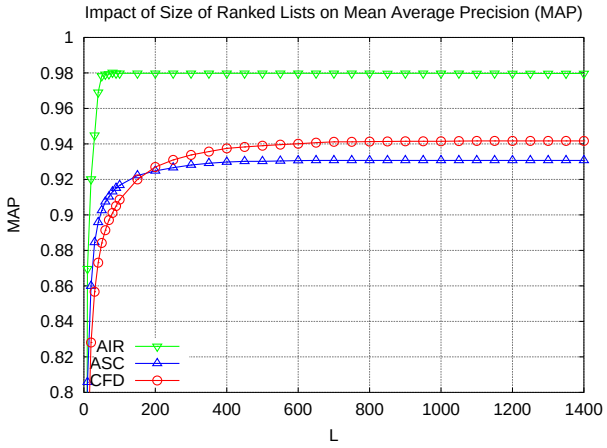

Figure 4: Impact of size L of ranked lists.

9.3. Adaptive Neighborhood Size

An experiment was conducted for analyzing the scores involved in the adaptive neighborhood size. The MPEG-7 [36] dataset and the CFD [47] shape descriptor were considered. Figure 5 presents the evolution of the reciprocal density score according to different values of k. The values of the first and second derivatives used for detecting the neighborhood size are also presented. All values are normalized in the interval [0,1]. The value of k is determined according to the second derivative, when it reaches a value equal to or less than zero. We can observe in Figure 5 that the second derivative (in red) reaches the x-axis when the reciprocal density score (in green) becomes stable.

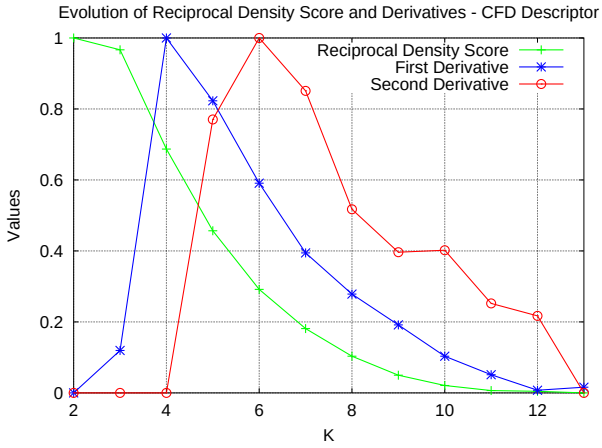

Figure 5: Adaptive-k analysis.

9.4. Shape, Color, and Texture Retrieval

The proposed method is evaluated in general image retrieval tasks considering shape, color, and texture properties. Three different datasets and twelve descriptors are considered. Experiments were conducted evaluating both the manual parameter setting and the adaptive neighborhood size approaches. We also consider different values of L ($L \leq 400$), and the whole ranked list. Statistical paired t-tests were conducted comparing the results before and after the use of the proposed algorithm.

Tables 3 and 4 present the experimental results based on Manual and Adaptive Neighborhood, respectively.

Table 3: Rank Diffusion based on Manual Neighborhood for general image retrieval tasks, considering shape, color, and texture features.

Descriptor	**Dataset**	**Original MAP**	**Rank Diff.**	**Gain**	**Stat. Sig. 99%**	**Rank Diffusion**	**Gain**	**Stat. Sig. 99%**
			Manual-k, $L = 400$			Manual-k, Full L		
Shape Descriptors								
SS [44]	MPEG-7	37.67%	52.17%	+38.49%	•	53.01%	+40.72%	•
BAS [45]	MPEG-7	71.52%	82.36%	+15.16%	•	83.03%	+16.09%	•
IDSC [46]	MPEG-7	81.70%	90.89%	+11.25%	•	91.09%	+11.49%	•
CFD [47]	MPEG-7	80.71%	93.75%	+16.16%	•	94.17%	+16.68%	•
ASC [48]	MPEG-7	85.28%	92.98%	+9.03%	•	93.07%	+9.13%	•
AIR [49]	MPEG-7	89.39%	97.98%	+9.61%	•	97.97%	+9.60%	•
Color Descriptors								
GCH [41]	Soccer	32.24%	36.34%	+12.72%	•	36.34%	+12.72%	•
ACC [42]	Soccer	37.23%	49.58%	+33.17%	•	49.58%	+33.17%	•
BIC [43]	Soccer	39.26%	50.39%	+28.35%	•	50.39%	+28.35%	•
Texture Descriptors								
LBP [50]	Brodatz	48.40%	50.42%	+4.17%	•	50.40%	+4.13%	•
CCOM [51]	Brodatz	57.57%	66.30%	+15.16%	•	66.46%	+15.44%	•
LAS [52]	Brodatz	75.15%	80.59%	+7.24%	•	80.68%	+7.36%	•

Table 4: Rank Diffusion based on Adaptive Neighborhood for general image retrieval tasks, considering shape, color, and texture features.

Descriptor	**Dataset**	**Original MAP**	**Rank Diff.**	**Gain**	**Stat. Sig. 99%**	**Rank Diffusion**	**Gain**	**Stat. Sig. 99%**
			Adaptive-k, $L = 400$			Adaptive-k, Full L		
Shape Descriptors								
SS [44]	MPEG-7	37.67%	39.94%	+6.03%	•	21.24%	-43.61%	•
BAS [45]	MPEG-7	71.52%	80.35%	+12.35%	•	81.89%	+14.50%	•
IDSC [46]	MPEG-7	81.70%	90.08%	+10.26%	•	90.31%	+10.54%	•
CFD [47]	MPEG-7	80.71%	93.86%	+16.29%	•	94.52%	+17.11%	•
ASC [48]	MPEG-7	85.28%	91.91%	+7.77%	•	91.99%	+7.87%	•
AIR [49]	MPEG-7	89.39%	97.74%	+9.34%	•	97.39%	+8.95%	•
Color Descriptors								
GCH [41]	Soccer	32.24%	34.70%	+7.63%	•	34.70%	+7.63%	•
ACC [42]	Soccer	37.23%	48.42%	+30.06%	•	48.42%	+30.06%	•
BIC [43]	Soccer	39.26%	49.57%	+26.26%	•	49.57%	+26.26%	•
Texture Descriptors								
LBP [50]	Brodatz	48.40%	48.28%	-0.25%		48.27%	-0.27%	
CCOM [51]	Brodatz	57.57%	65.10%	+13.08%	•	65.24%	+13.32%	•
LAS [52]	Brodatz	75.15%	79.02%	+5.15%	•	79.17%	+5.35%	•

Positive gains with statistical significance can be observed for all descriptors and datasets for the manual neighborhood, reaching high effectiveness gains up to +40.72%. For the adaptive neighborhood, the results are posi-

tive for most of descriptors, except for SS [44] and LBP [50] descriptors. The effectiveness results obtained for partial ($L \leq 400$) and the entire ranked lists (full L) are very similar, demonstrating that only a sub-set of ranked lists is enough to obtain high effectiveness gains. For most descriptors, the effectiveness gains of manual and adaptive neighborhood are also similar, with a narrow margin for the manual setting.

A visual example of the impact of the algorithm on retrieval results is illustrated in Figure 6.

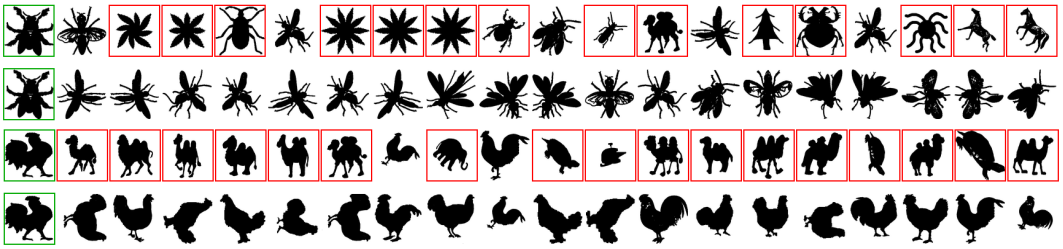

Figure 6: Visual examples of the effectiveness gains considering the CFD [47] and ASC [48] descriptor on the MPEG-7 [36] dataset. Retrieval results before and after the use of the algorithm: query image with green border and wrong images with red borders.

Table 5 presents the results for rank aggregation tasks, considering two descriptors of each visual property. As it can be observed, high effective results are also obtained.

9.5. Object Retrieval

We evaluated the proposed method in object retrieval tasks considering the ETH-80 [39]. Table 6 presents the effectiveness results obtained by four color descriptors before and after the use of the algorithm.

Positive effectiveness gains can be observed for all descriptors, considering the both manual and adaptive neighborhood size selection. A relative gain of +8.67% was reached considering the CSD [53] descriptor.

9.6. Natural Image Retrieval

We evaluate the proposed method in natural image retrieval tasks considering two popular datasets: the University of Kentucky Recognition Benchmark - UKBench [40] and the Holidays [35] dataset.

Table 7 presents the effectiveness results for the UKBench [40] dataset. The N-S score is used as effectiveness measure, varying between 1 and 4.

Table 5: Rank Diffusion for rank aggregation tasks (shape, color, and texture).

Descriptor	**Type**	**Dataset**	**Neighbor Selection**	**Score (MAP)**
CFD [47]	Shape	MPEG-7	-	80.71%
ASC [48]	Shape	MPEG-7	-	85.28%
CFD+ASC	Shape	MPEG-7	Manual-k	99.29%
CFD+ASC	Shape	MPEG-7	Adaptive-k	99.00%
ACC [42]	Color	Soccer	-	37.23%
BIC [43]	Color	Soccer	-	39.26%
BIC+ACC	Color	Soccer	Manual-k	49.75%
BIC+ACC	Color	Soccer	Adaptive-k	48.20%
CCOM [51]	Texture	Brodatz	-	57.57%
LAS [52]	Texture	Brodatz	-	75.15%
LAS+CCOM	Texture	Brodatz	Manual-k	84.52%
LAS+CCOM	Texture	Brodatz	Adaptive-k	82.88%

Table 6: Rank Diffusion for object retrieval on ETH-80 [39] dataset.

Descriptor	**Original Score (MAP)**	**Rank Diffusion** *Manual-k*	**Rank Diffusion** *Adaptive-k*
BIC [43]	49.72%	51.99%	52.66%
ACC [42]	48.50%	51.49%	51.99%
CSD [53]	48.46%	51.30%	52.66%
GCH [41]	41.62%	42.40%	43.42%

This score corresponds to the number of relevant images among the first four image returned (the highest achievable score is 4). We can observe significant improvements for N-S scores. Notice, for example, the Caffe [58] convolutional neural network, which is improved from 3.31 to 3.61. The results are even more impressive considering the rank aggregation tasks, reaching a N-S score of 3.94.

Table 8 presents the results obtained by the Rank Diffusion method on the Holidays [35] dataset. Significant effectiveness gains can be observed for all features, reaching +13.4% for the CNN-Caffe [58] feature.

9.7. Comparison with Other Approaches

The proposed method is also evaluated in comparison with various other state-of-the-art approaches and recently proposed retrieval approaches. Experiments were conducted on three image datasets: MPEG-7 [36], Holi-

Table 7: Rank Diffusion for natural image retrieval on the UKBench [40] dataset.

Descriptor	**Original N-S Score**	**Rank Diffusion**	**Rank Diffusion**
		Manual-k	*Adaptive-k*
CEED-SPy [56, 57]	2.81	3.10	3.10
FCTH-SPy [60, 57]	2.91	3.19	3.19
SCD [55]	3.15	3.35	3.36
ACC-SPy [42, 57]	3.25	3.51	3.51
CNN-Caffe [58]	3.31	3.61	3.61
ACC [57]	3.36	3.60	3.60
VOC [61]	3.54	3.72	3.72
VOC+ACC	-	3.90	3.90
VOC+CNN-Caffe	-	3.90	3.90
ACC+CNN-Caffe	-	3.87	3.88
VOC+ACC+CNN-Caffe	-	**3.94**	**3.94**

Table 8: Rank Diffusion for natural image retrieval on the Holidays [35] dataset.

Descriptor	**Original MAP**	**Rank Diffusion**	**Relative Gain**
JCD [54]	52.83%	54.03%	+2.27%
SCD [55]	54.26%	57.11%	+5.25%
CEED-SPy [56, 57]	56.09%	58.06%	+3.51%
ACC [42]	64.29%	71.48%	+11.18%
CNN-Caffe [58]	64.09%	72.77%	+13.54%
CNN-OverFeat [59]	82.59%	**86.17%**	+4.33%
ACC + CEED-SPy	-	71.59%	+11.35%
ACC + Caffe + OverFeat	-	85.93%	+4.04%

days [35] and UKBench [40], which are popular datasets commonly used as benchmark for image retrieval and post-processing methods.

Table 9: Comparison with state-of-the-art on the MPEG-7 [36] dataset.

Shape Descriptors		
DDGM [62]	-	80.03%
CFD [47]	-	84.43%
IDSC [46]	-	85.40%
SC [63]	-	86.80%
ASC [48]	-	88.39%
AIR [49]	-	93.67%
Post-Processing Methods		
Algorithm	**Descriptor(s)**	**Score**
Graph Transduction [7]	IDSC	91.61%
Shortest Path Propagation [64]	IDSC	93.35%
Mutual kNN Graph [8]	IDSC	93.40%
Locally C. Diffusion Process [6]	ASC	95.96%
RL-Sim [11]	CFD	95.33%
Pairwise Recommendation [65]	CFD	96.15%
Rank Diffusion	**CFD**	**96.19%**
Tensor Product Graph [5]	ASC	96.47%
Self-Smoothing Operator [9]	SC+IDSC	97.64%
Co-Transduction [66]	SC+IDSC	97.72%
SCA [14]	SC+IDSC	99.01%
Tensor Product Graph [5]	AIR	99.99%
Generic Diffusion Process [24]	AIR	100%
Neighbor Set Similarity [13]	AIR	100%
Rank Diffusion	**AIR**	**100%**

Table 9 presents the obtained results on the MPEG-7 [36] dataset in comparison with various other state-of-the-art post-processing methods. The bull's eye score, which counts the matching shapes within the top-40 ranked images, is used as evaluation measure. As it can be observed, the effectiveness results of the proposed method compares favorably with the recent retrieval approaches.

Table 10 shows the MAP scores obtained on the Holidays [35] dataset. The effectiveness results are comparable to most of considered methods. Finally, Table 11 presents the results of Rank Diffusion method on the UKBench [40] dataset in comparison with recent retrieval approaches. The Rank Diffusion method yielded the best N-S score in comparison with other state-of-the-art methods.

Table 10: Comparison with state-of-the-art on the Holidays [35] dataset.

MAP scores for recent retrieval methods.			
Jégou *et al.* [35]	Li *et al.* [67]	Zheng *et al.* [68]	Tolias *et al.* [69]
75.07%	89.20%	85.80%	82.20%

Qin *et al.* [70]	Zheng *et al.* [71]	**Rank Diffusion** CNN-OverFeat
84.40%	85.20%	**86.17%**

Table 11: Comparison with state-of-the-art on the UKBench [40] dataset.

N-S scores for recent retrieval methods					
Zheng *et al.* [72]	Qin *et al.* [26]	Wang *et al.* [73]	Zhang *et al.* [74]	Zheng *et al.* [34]	Bai *et al.* [14]
3.57	3.67	3.68	3.83	3.84	3.86

Xie *et al.* [75]	Rank Diffusion VOC+CNN	Rank Diffusion VOC+ACC	**Rank Diffusion** VOC+ACC+CNN
3.89	3.90	3.90	**3.94**

10. Conclusions

Post-processing procedures based on unsupervised learning approaches have been establishing as an indispensable tool for improving the effectiveness of CBIR systems. Since diffusion process methods require high computational efforts, rank-based approaches attracted a lot of research attention recently to circumvent their limitations. In this paper, a novel rank diffusion method is proposed exploiting characteristics of both diffusion and rank-based approaches. Despite the use of a diffusion strategy, a low-complexity re-ranking algorithm is proposed, once only rank information is considered. A rigorous experimental evaluation demonstrated the effectiveness of the proposed approach.

Future work focuses on the deep investigation of contextual information encoded in the rank similarity matrix. We also plan to investigate the use of indexing schemes to speed up processing time even further.

Acknowledgments

The authors are grateful to São Paulo Research Foundation – FAPESP (grants #2013/08645-0 and #2013/50169-1), CNPq (grants #306580/2012-8 and #484254/2012-0), CAPES, AMD, and Microsoft Research.

References

[1] M. S. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: State of the art and challenges, ACM Transactions on Multimedia Computing, Communications, and Applications 2 (1) (2006) 1–19.

[2] Y. Liu, D. Zhang, G. Lu, W.-Y. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recognition 40 (1) (2007) 262 – 282.

[3] B. Thomee, M. Lew, Interactive search in image retrieval: a survey, International Journal of Multimedia Information Retrieval 1 (2) (2012) 71–86.

[4] C. D. Ferreira, J. A. dos Santos, R. da S. Torres, M. A. Gonçalves, R. C. Rezende, W. Fan, Relevance feedback based on genetic programming for image retrieval, Pattern Recogninion Letters 32 (1) (2011) 27–37.

[5] X. Yang, L. Prasad, L. Latecki, Affinity learning with diffusion on tensor product graph, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 28–38.

[6] X. Yang, S. Koknar-Tezel, L. J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval., in: CVPR, 2009, pp. 357–364.

[7] X. Bai, X. Yang, L. J. Latecki, W. Liu, Z. Tu, Learning context-sensitive shape similarity by graph transduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (5) (2010) 861–874.

[8] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: ACCV, 2009, pp. 655–666.

[9] J. Jiang, B. Wang, Z. Tu, Unsupervised metric learning by self-smoothing operator, in: ICCV, 2011, pp. 794–801.

[10] D. C. G. Pedronette, J. Almeida, R. da S. Torres, A scalable re-ranking method for content-based image retrieval, Information Sciences 265 (1) (2014) 91–104.

[11] D. C. G. Pedronette, R. da S. Torres, Image re-ranking and rank aggregation based on similarity of ranked lists, Pattern Recognition 46 (8) (2013) 2350–2360.

[12] D. C. G. Pedronette, R. da S. Torres, Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks, Image and Vision Computing 32 (2) (2014) 120–130.

[13] X. Bai, S. Bai, X. Wang, Beyond diffusion process: Neighbor set similarity for fast re-ranking, Information Sciences 325 (2015) 342 – 354.

[14] S. Bai, X. Bai, Sparse contextual activation for efficient visual re-ranking, IEEE Transactions on Image Processing 25 (3) (2016) 1056–1069.

[15] D. C. G. Pedronette, R. da S. Torres, Rank diffusion for context-based image retrieval, in: ACM International Conference on Multimedia Retrieval (ICMR'2016), 2016.

[16] R. T. Calumby, M. A. Gonçalves, R. da Silva Torres, On interactive learning-to-rank for ir: Overview, recent advances, challenges, and directions, Neurocomputing 208 (2016) 3 – 24.

[17] W. Chen, T. Liu, Y. Lan, Z. Ma, H. Li, Ranking measures and loss functions in learning to rank, in: Conference on Neural Information Processing Systems (NIPS), 2009, pp. 315–323.

[18] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, H. Li, Learning to rank: From pairwise approach to listwise approach, in: International Conference on Machine Learning, ICML '07, 2007, pp. 129–136.

[19] E. de Ves, X. Benavent, I. Coma, G. Ayala, A novel dynamic multi-model relevance feedback procedure for content-based image retrieval, Neurocomputing 208 (2016) 99 – 107.

[20] Z. Ghahramani, Unsupervised Learning. Advanced Lectures on Machine Learning, 2004, pp. 72–112.

[21] C. Qin, S. Song, G. Huang, L. Zhu, Unsupervised neighborhood component analysis for clustering, Neurocomputing 168 (2015) 609 – 617.

[22] R. Hu, X. Zhu, D. Cheng, W. He, Y. Yan, J. Song, S. Zhang, Graph self-representation method for unsupervised feature selection, Neurocomputing 220 (2017) 130 – 137.

[23] H. Jegou, C. Schmid, H. Harzallah, J. Verbeek, Accurate image search using the contextual dissimilarity measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (1) (2010) 2–11.

[24] M. Donoser, H. Bischof, Diffusion Processes for Retrieval Revisited, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1320–1327.

[25] Y. Chen, X. Li, A. Dick, R. Hill, Ranking consistency for image matching and object retrieval, Pattern Recognition 47 (3) (2014) 1349 – 1360.

[26] D. Qin, S. Gammeter, L. Bossard, T. Quack, L. van Gool, Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors, in: CVPR, 2011, pp. 777 –784.

[27] B. A. Stickler, E. Schachinger, The Random Walk and Diffusion Theory, Cham, 2016, pp. 271–295.

[28] O. C. Ibe, Elements of Random Walk and Diffusion Processes, 1st Edition, Wiley Publishing, 2013.

[29] R. R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5 – 30.

[30] X. Shen, Z. Lin, J. Brandt, S. Avidan, Y. Wu, Object retrieval and localization with spatially-constrained similarity measure and k-nn reranking, in: CVPR, 2012, pp. 3013 –3020.

[31] R. da S. Torres, A. X. Falcão, Content-Based Image Retrieval: Theory and Applications, Revista de Informática Teórica e Aplicada 13 (2) (2006) 161–185.

[32] J. Almeida, R. da S. Torres, N. J. Leite, BP-tree: An efficient index for similarity search in high-dimensional metric spaces, in: CIKM, 2010, pp. 1365–1368.

[33] R. Fagin, R. Kumar, D. Sivakumar, Comparing top k lists, in: ACM-SIAM Symposium on Discrete algorithms (SODA'03), 2003, pp. 28–36.

[34] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, Q. Tian, Query-adaptive late fusion for image search and person re-identification, in: CVPR'2015, 2015.

[35] H. Jegou, M. Douze, C. Schmid, Hamming embedding and weak geometric consistency for large scale image search, in: European Conference on Computer Vision, ECCV '08, 2008, pp. 304–317.

[36] L. J. Latecki, R. Lakmper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: CVPR, 2000, pp. 424–429.

[37] J. van de Weijer, C. Schmid, Coloring local feature extraction, in: ECCV, 2006, pp. 334–348.

[38] P. Brodatz, Textures: A Photographic Album for Artists and Designers, Dover, 1966.

[39] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, in: CVPR, Vol. 2, 2003, pp. II–409–15 vol.2.

[40] D. Nistér, H. Stewénius, Scalable recognition with a vocabulary tree, in: CVPR, Vol. 2, 2006, pp. 2161–2168.

[41] M. J. Swain, D. H. Ballard, Color indexing, International Journal on Computer Vision 7 (1) (1991) 11–32.

[42] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlograms, in: CVPR, 1997, pp. 762–768.

[43] R. O. Stehling, M. A. Nascimento, A. X. Falcão, A compact and efficient image retrieval approach based on border/interior pixel classification, in: CIKM, 2002, pp. 102–109.

[44] R. da S. Torres, A. X. Falcão, Contour Salience Descriptors for Effective Image Retrieval and Analysis, Image and Vision Computing 25 (1) (2007) 3–13.

[45] N. Arica, F. T. Y. Vural, BAS: a perceptual shape descriptor based on the beam angle statistics, Pattern Recognition Letters 24 (9-10) (2003) 1627–1639. doi:http://dx.doi.org/10.1016/S0167-8655(03)00002-3.

[46] H. Ling, D. W. Jacobs, Shape classification using the inner-distance, IEEE TPAMI 29 (2) (2007) 286–299. doi:http://dx.doi.org/10.1109/TPAMI.2007.41.

[47] D. C. G. Pedronette, R. da S. Torres, Shape retrieval using contour features and distance optmization, in: VISAPP, Vol. 1, 2010, pp. 197 – 202.

[48] H. Ling, X. Yang, L. J. Latecki, Balancing deformability and discriminability for shape matching, in: ECCV, Vol. 3, 2010, pp. 411–424.

[49] R. Gopalan, P. Turaga, R. Chellappa, Articulation-invariant representation of non-planar shapes, in: ECCV, Vol. 3, 2010, pp. 286–299.

[50] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE TPAMI 24 (7) (2002) 971–987.

[51] V. Kovalev, S. Volmer, Color co-occurence descriptors for querying-by-example, in: ICMM, 1998, p. 32.

[52] B. Tao, B. W. Dickinson, Texture recognition and image retrieval using gradient indexing, JVCIR 11 (3) (2000) 327–342.

[53] B. Manjunath, J.-R. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11 (6) (2001) 703–715.

[54] K. Zagoris, S. Chatzichristofis, N. Papamarkos, Y. Boutalis, Automatic image annotation and retrieval using the joint composite descriptor, in: PCI, 2010, pp. 143–147.

[55] B. Manjunath, J.-R. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11 (6) (2001) 703–715.

[56] S. A. Chatzichristofis, Y. S. Boutalis, Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval, in: ICVS, 2008, pp. 312–322.

[57] M. Lux, Content based image retrieval with LIRe, in: ACM Multimedia '11, 2011.

[58] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093.

[59] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'14), pp. 512–519.

[60] S. A. Chatzichristofis, Y. S. Boutalis, Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval, in: WIAMIS, 2008, pp. 191–196.

[61] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, T. Han, Contextual weighting for vocabulary tree based image retrieval, in: ICCV'2011, 2011, pp. 209–216.

[62] Z. Tu, A. L. Yuille, Shape matching and recognition - using generative models and informative features, in: ECCV, 2004, pp. 195–209.

[63] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE TPAMI 24 (4) (2002) 509–522.

[64] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, N. Tang, Learning context-sensitive similarity by shortest path propagation, Pattern Recognition 44 (10-11) (2011) 2367–2374.

[65] D. C. G. Pedronette, R. da S. Torres, Exploiting pairwise recommendation and clustering strategies for image re-ranking, Information Sciences 207 (2012) 19–34.

[66] X. Bai, B. Wang, C. Yao, W. Liu, Z. Tu, Co-transduction for shape retrieval, IEEE Transactions on Image Processing 21 (5) (2012) 2747–2757.

[67] X. Li, M. Larson, A. Hanjalic, Pairwise geometric matching for large-scale object retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2015), 2015, pp. 5153–5161.

[68] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: Coupled multi-index for accurate image retrieval, in: CVPR'2014, 2014, pp. 1947–1954.

[69] G. Tolias, Y. Avrithis, H. Jgou, To aggregate or not to aggregate: Selective match kernels for image search, in: IEEE International Conference on Computer Vision (ICCV'2013), 2013, pp. 1401–1408.

[70] D. Qin, C. Wengert, L. V. Gool, Query adaptive similarity for large scale object retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2013), 2013, pp. 1610–1617.

[71] L. Zheng, S. Wang, Q. Tian, Coupled binary embedding for large-scale image retrieval, IEEE Transaction on Image Processing 23 (8) (2014) 3368–3380.

[72] L. Zheng, S. Wang, Q. Tian, Lp-norm IDF for scalable image retrieval, IEEE Transactions on Image Processing 23 (8) (2014) 3604–3617.

[73] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, Z. Tu, Unsupervised metric fusion by cross diffusion, in: CVPR, 2012, pp. 3013 –3020.

[74] S. Zhang, M. Yang, T. Cour, K. Yu, D. Metaxas, Query specific rank fusion for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (4) (2015) 803–815.

[75] L. Xie, R. Hong, B. Zhang, Q. Tian, Image classification and retrieval are one, in: ACM ICMR '15, 2015, pp. 3–10.