Pattern Recognition 75 (2018) 161–174

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks

Daniel Carlos Guimarães Pedronette*, Filipe Marcel Fernandes Gonçalves, Ivan Rizzo Guilherme

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

ARTICLE INFO

Article history:
Received 8 November 2016
Revised 10 March 2017
Accepted 13 May 2017
Available online 13 May 2017

Keywords:
Content-based image retrieval
Unsupervised manifold learning
Reciprocal kNN graph
Connected components

ABSTRACT

Performing effective image retrieval tasks, capable of exploiting the underlying structure of datasets still constitutes a challenge research scenario. This paper proposes a novel manifold learning approach that exploits the intrinsic dataset geometry for improving the effectiveness of image retrieval tasks. The underlying dataset manifold is modeled and analyzed in terms of a Reciprocal kNN Graph and its Connected Components. The method computes the new retrieval results on an unsupervised way, without the need of any user intervention. A large experimental evaluation was conducted, considering different image retrieval tasks, various datasets and features. The proposed method yields better effectiveness results than various methods recently proposed, achieving effectiveness gains up to +40.75%.

1. Introduction

The fast and continuous growth of image collections in different application domains has demanded the development of effective and efficient methods for retrieving the images based on their visual content. Content-Based Image Retrieval (CBIR) [1] has established in this scenario as a solid solution, attracting increasing interest of academia and industry. The development of CBIR systems was firstly supported by sophisticated low-level feature extraction algorithms, giving rise to a broad range of image features, from the traditional global features [2] (based on shape, color, and texture properties) to the most recent convolution-neural-network-based features [3,4].

Despite of the continuous and consistent development of visual features in the last decades [2,5,6], effectively measuring the similarity among images remains a challenging problem in image classification and retrieval tasks. The vast majority of image features, including the most recent, commonly measure the distance or similarity between images based on pairwise comparison of feature vectors. However, the classical pairwise distance measures, as the Euclidean distance, often fail to produce effective results in various scenarios. Additionally, the need for appropriate ways to measure the distance or similarity between complex data is ubiquitous in many machine learning, pattern recognition and retrieval applications [7,8].

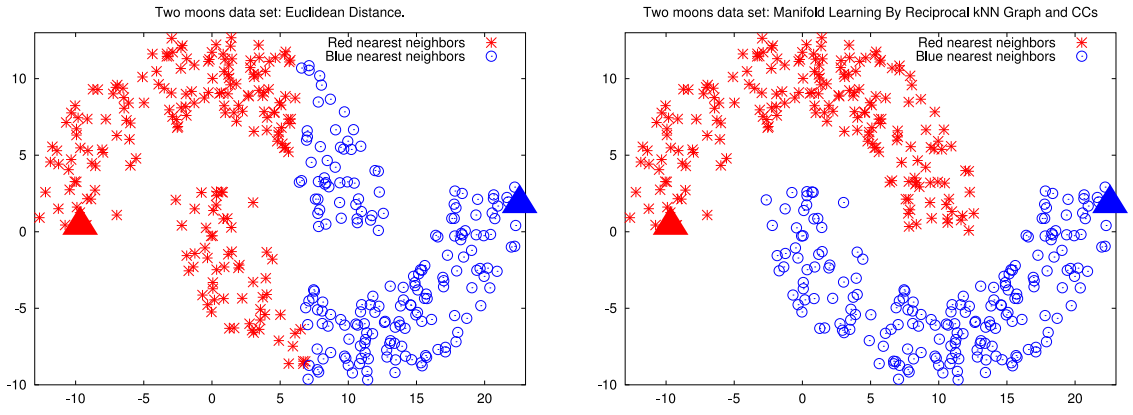
Mainly due to this difficulties, the research focus was gradually shifted from designing low-level features to aspects related to higher level aspects [9]. Based on this assumption, the use of machine learning methods was quickly spread in order to associate low-level features with high-level query concepts. In image retrieval applications, for example, several relevance feedback approaches [10–13] have been proposed. Such approaches obtain supervised information through user interactions with the aim of learning new distance measures capable of encoding user preferences.

In this scenario, distance metric learning methods have been demonstrating a great potential and can be considered as a promising solution [7,8,14–16]. Metric learning can be broadly defined as the transformation of data samples from the original space to another feature space by reducing the intra-class variation and increasing the inter-class variations [15]. Most of approaches are supervised, aiming at learning a distance metric from a number of training samples with side information, i.e., relevance judgments defined by pairwise constraints obtained from the users [8]. Motivated by the sparsity of training information, semi-supervised learning approaches [13,17] were also proposed, incorporating unlabeled data in the distance metric learning procedure [18].

However, in certain retrieval scenarios, the training data is nonexistent or infeasible to obtain, which leaded to the development of unsupervised approaches. In this way, various approachess [19–25] have been put efforts in post-processing the retrieval results by learning a new distance on an unsupervised way. Overall, such methods aims at improving the distance measures in

* Corresponding author.
E-mail addresses: pedronette@gmail.com, daniel@rc.unesp.br (D.C.G. Pedronette).

http://dx.doi.org/10.1016/j.patcog.2017.05.009
0031-3203/



(a) Euclidean distance. (b) Reciprocal kNN Graph + CCs.

Fig. 1. Two-Moons dataset: the color of the points is assigned according to the lower distance to the highlighted query samples (triangles). Distances computed by the **Euclidean distance** in the left **(a)**, and using the **Reciprocal kNN Graph + CCs** in the right **(b)**, which properly considers the dataset manifold.

CBIR systems without the need of any user intervention. The central idea consists in replacing pairwise distance measures by more global measures [20], capable of considering the dataset structure.

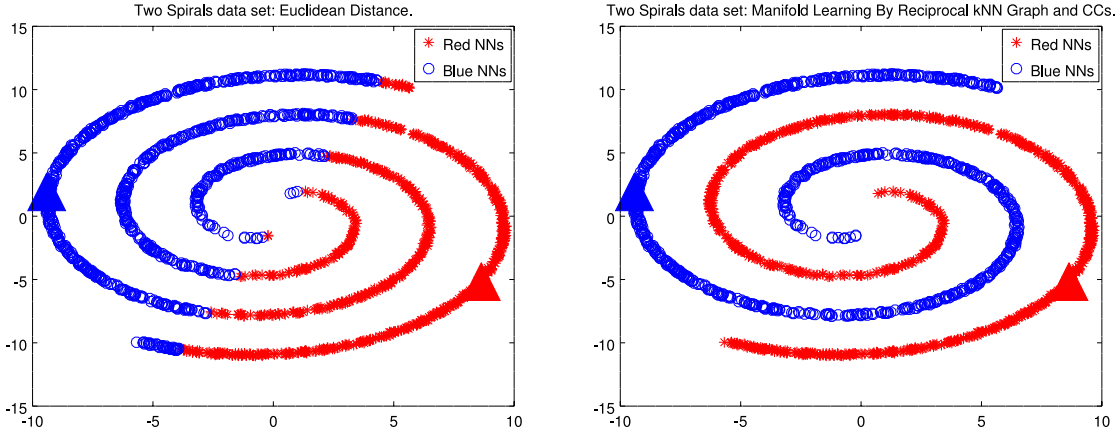
Various unsupervised approaches are based on diffusion process [20–22,26] and graph learning [27] methods. Other unsupervised learning methods [25,28] have focused on exploiting the dataset manifold, mainly due to the incapacity of pairwise distances for considering the dataset structure. In fact, the analysis of dataset manifold have emerged as a promising tool in different learning scenarios [29–32]. For retrieval tasks, once images and multimedia objects are often modeled as high dimensional points in an Euclidean space and they often live in a much lower-dimensional intrinsic space, discovering and exploiting the intrinsic manifold structure constitutes a central problem [22]. More recently, rank-based approaches also have been attracted a lot of attention [33,34], since the use of ranking information provide relevant advantages, as lowers computational efforts and independence of distance measures. In addition, such unsupervised approaches can be used for combining information from different retrieval models, as Bag-of-Words (BoW) and Convolutional Neural Networks (CNN), which have achieving high effective results recently [35–37].

In this paper, a novel unsupervised manifold learning algorithm for image retrieval is proposed by exploiting the Reciprocal *k*NN Graph and its Connected Components (CCs). The algorithm models the dataset similarity structure in a graph, based on the reciprocal references encoded in the ranking information. The Reciprocal *k*NN Graph is constructed at different depths of *k*, providing a multi-level analysis. The reciprocity relationships are used for identifying more reliable similarity information. In this way, while the edges of the Reciprocal *k*NN Graph provide a strong indication of similarity, the Connected Components are exploited for capturing the intrinsic geometry of the dataset manifold. The information encoded in the graph and Connected Components is used to improve the effectiveness of distance measures, giving rise a new and more effective retrieval results.

The capacity of exploiting the geometry of the dataset manifold for computing a new distance can be observed in Figs. 1 and 2. In Fig. 1, a query sample is selected in each moon, represented by a labeled point marked with a triangle. The color of other points are determined according to the closest labeled point. Fig. 1a shows a representation of the *Two-Moons* dataset considering the Euclidean distance. Once the geometry of the dataset is not considered, a large number points are misclassified. Fig. 1b illustrates the color classification computed by the proposed manifold learning algorithm. As we can observe, a perfect classification is produced, once the whole geometry of the dataset is exploited. Fig. 2 presents an analogous problem, considering the more challenger *Two-Spirals* dataset. Again, the Euclidean distance fail to classify most of points (Fig. 2a), while the proposed manifold learning respect the dataset manifold, producing a perfect classification (Fig. 2b).

The main contributions of the proposed approach in face of the related work are highlighted as follows:

- Although the reciprocal references have been broadly exploited last years [23,24,38–40], the proposed algorithm mainly differs from the other approaches regarding the use of Connected Components and the graph construction at different depths of the *k*-neighborhood, which enables a more gradual analysis at different levels of similarity. More specifically regarding [23,38], we can emphasize:
  - The Graph Fusion [23] also takes into account the *k*-reciprocal neighborhood for building the graph, but requires the computation of the Jaccard measure for assigning weight to edges, while the proposed method uses only information of reciprocal neighborhood. The Graph Fusion [23] performs a ranking step using a transition matrix based on PageRank or a greedy algorithm, while our proposal exploits the Connected Components;
  - The reciprocal neighborhood is also exploited in [38], which uses a directed graph, while the propose method uses a weighted undirected graph. Two different neighborhood sizes (k and k_{max}) are required in [38], separating different parts of the ranked retrieval list with different distance measures. In opposite, the proposed method considers only one neighborhood size, computing an uniform single measure.
- The proposed algorithm requires lower computational efforts than diffusion-based approaches [20–22,26]. Such methods compute successive powers of affinity matrices, while the connectivity of the proposed graph is defined in terms of only top-k positions. On the other hand, when compared to other rank-based approaches [33,34], the proposed method provides a geometric and more intuitive interpretation of the data;
- The proposed algorithm is related to the manifold learning method based on the correlation graph [25], which defines the graph connectivity using different levels of correlation measures and exploits strongly Connected Components. In contrast, the proposed method provides a simpler and more efficient formulation, in terms of reciprocal references and Connected Components.



(a) Euclidean distance. (b) Reciprocal kNN Graph + CCs.

Fig. 2. Two-Spirals dataset: the color of the points is assigned according to the lower distance to the highlighted query samples (triangles). Distances computed by the **Euclidean distance** in the left **(a)**, and using the **Reciprocal kNN Graph + CCs** in the right **(b)**, which properly considers the dataset manifold.

The proposed algorithm was evaluated through a large experimental evaluation, considering various distinct image retrieval scenarios. Experiments were conducted on 6 public datasets and considering 21 different image descriptors, including global (shape, color, and texture), local, and convolution-neural-network-based features. Experimental results demonstrates that significant effectiveness gains can be obtained in various image retrieval tasks. The proposed algorithm achieves effectiveness gains up to +40.75%, yielding better or comparable effectiveness results than various state-of-the-art and recent related approaches.

The remaining of the paper is organized as follows: Section 2 discusses the problem formulation and Section 3 presents the unsupervised manifold learning algorithm. Section 4 describes the experimental evaluation and, finally, Section 5 draws the conclusions and discusses future work.

2. Image retrieval and rank model

This section presents a formal definition of the image retrieval and ranking model considered along the paper. Let $\mathcal{C} = \{img_1, img_2, \ldots, img_n\}$ be an image collection, where n denotes the size of the collection. Let $\mathcal{D}$ be an image descriptor defined [41] as a tuple $\mathcal{D} = (\epsilon, \rho)$, where:

- ϵ: $\hat{I} \rightarrow \mathbb{R}^n$ is a function, which extracts a feature vector $v_{\hat{I}}$ from an image $\hat{I}$; and
- 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.

Therefore, a distance between two images img_i, img_j can be computed by $\rho(\epsilon(img_i), \epsilon(img_j))$. For readability purposes, the notation $\rho(i, j)$ is used along the paper to refer to the distance between images img_i and img_j. Notice that, although the model is defined in terms of images, small changes can become it suitable to many recognition problems.

Based on the distance function ρ, a retrieval and ranking model can be derived. For a general image retrieval task, a ranked list τ_q can be computed in response to a query image img_q, according to the distance function ρ. The top positions of ranked lists are expected to contain the most relevant images with regard to the query image, such that only the top-L ranked images are considered, with $L \ll n$.

The ranked list τ_q can be formally defined as a permutation $(img_1, img_2, \ldots, img_L)$ of the subset $\mathcal{C}_L \subset \mathcal{C}$, which contains the L most similar images to query image img_q, such that and $|\mathcal{C}_L| = L$. A permutation τ_q is a bijection from the set $\mathcal{C}_L$ onto the set $[n_L] = \{1, 2, \ldots, L\}$. For a permutation τ_q, we interpret $\tau_q(i)$ as the position (or rank) of image img_i in the ranked list τ_q. If img_i is ranked before img_j in the ranked list of img_q ($\tau_q(i) < \tau_q(j)$), then $\rho(q, i) \leq \rho(q, j)$.

Considering every image in the collection as a query image, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained. The set $\mathcal{T}$ represents a rich source of distance/similarity information about the collection $\mathcal{C}$, once the ranked lists contain the most similar images in priority order.

The *Reciprocal kNN Graph and Connected Components* algorithm discussed in this paper aims at exploiting the ranking information encoded in the set $\mathcal{T}$ for computing a new and more effective distance function ρ_r, and therefore improving the effectiveness of image retrieval tasks.

3. Reciprocal kNN graph and Connected Components for unsupervised manifold learning

The proposed manifold learning algorithm aims at capturing the underlying dataset structure by exploiting and analyzing its ranking information. In opposite to distance measures, which compare only pairs of images, ranked lists establish a deeper relationship, involving the comparison of the query image with all dataset images [34]. The ranked lists constitute a rich source of similarity information, including the neighborhood set, which can be modeled in terms of top-k rank positions.

However, different from metric distance (or similarity) measures, the nearest neighbor relationships are not symmetric [38]. While $\rho(i, j) = \rho(j, i)$, the presence of the image img_j in the k-neighborhood of img_i does not imply that the img_j neighborhood contains img_i. Morever, studies [42] have shown that the improvement of nearest neighborhood relationships can enhance the retrieval effectiveness. Recently, the k-reciprocal neighborhood has been exploited in image retrieval [24,38–40], once it defines a symmetric and solid indication of similarity, specially for small values of k.

Although more reliable, the reciprocal neighborhood defines less relationships among images, restricting the similarity analysis. Given a Reciprocal kNN Graph, where the edges indicates a reciprocal neighborhood relationship, the graph is often sparse, providing less information about the underlying manifold. In this scenario, the Connected Components are exploited for expanding the

neighborhood through reliable edges and, at the same time, taking into account the dataset structure.

Based on these concepts, the proposed algorithm models the ranking information using a graph through four main steps:

1. **Rank normalization:** first, a Rank Normalization procedure is performed, updating the initial ranked lists according to the ranking references;
2. **Reciprocal kNN Graph:** the Reciprocal *k*NN Graph is constructed defining edges only between images with reciprocal references at top-*k* positions;
3. **Connected Components:** the Connected Components (CCs) are computed and used for identifying sets of similar images;
4. **Reciprocal kNN Graph Distance:** a new distance measure is computed based on similarity information encoded in the graph and CCs. Different depths of *k* are considered for the graph construction, defining different levels of confidence.

Each of the main steps of the algorithm are detailed and formally defined in next sub-sections.

3.1. Rank normalization

Once the rank relationships are not symmetric, an image img_i well ranked for a query img_j does not imply that img_j is well ranked for a query img_i. In this way, improving the symmetry of the *k*-neighborhood often benefits the effectiveness of image retrieval methods [42].

Therefore, a rank normalization approach is employed considering both mutual [34] and reciprocal [38] neighborhood. While the mutual neighborhood sums rank positions from both ranked lists, the reciprocal neighborhood considers only the maximum, and more reliable value. A combined rank normalized distance ρ_n is proposed, defined as:

$$\rho_n(i,j) = \tau_i(j) + \tau_j(i) + max(\tau_i(j), \tau_j(i)), \tag{1}$$

where $\tau_i(j) \leq L$. Based on the ρ_n distance, the ranked lists are updated until the top-*L* positions using a stable sorting algorithm.

3.2. Reciprocal kNN Graph

The *Reciprocal kNN Graph* can be defined as an undirected graph $G_r = (V, E)$, where the set of vertices V is given by the image collection $V = \mathcal{C}$ and each image is represented by a node. The edge set E is computed based on the *k*-reciprocal neighborhood considering different thresholds for *k*.

For determining the reciprocal neighborhood, we first define a neighborhood set. Given a query image img_q, a neighborhood set $\mathcal{N}(q,k)$ that contains the k most similar images to img_q can be defined as follows:

$$\mathcal{N}(q,k) = \{\mathcal{S} \subseteq \mathcal{C}, |\mathcal{S}| = k \wedge \forall img_i \in \mathcal{S}, \\ img_j \in \mathcal{C} - \mathcal{S} : \tau_q(i) < \tau_q(j)\}. \tag{2}$$

Once the nearest neighbor relationships are not symmetric [38,42], the set of *k*-reciprocal nearest neighbors of image img_q can be defined [38] as:

$$\mathcal{N}_r(q,k) = \{img_i \in \mathcal{N}(q,k) \wedge img_q \in \mathcal{N}(i,k)\}. \tag{3}$$

Let t_k denotes a threshold which defines the value of k at a given moment of algorithm execution, the edge set E can be formally defined as:

$$E = \{(img_q, img_j) \mid img_j \in \mathcal{N}_r(q,t_k)\}. \tag{4}$$

Therefore, we can interpret that there will be edge from img_q to img_j if the images are reciprocal neighbors until the t_k positions.

3.3. Connected Components

An effective context-based measure redefines the distance among images by exploiting more reliable similarity information encoded in the dataset. In this way, the reciprocal neighborhood provides a strong indication of similarity [38].

Although very precise, only a small number of edges is created, giving rise to a sparse and disconnected graph. However, the information of graph connectivity can be exploited for considering the geometry of the dataset manifold and expanding the similarity neighborhood [43]. The Connected Components (CC) of the Reciprocal kNN Graph are used with this objective, allowing a increase of similarity among images in the same CC. Therefore, the information encoded in the CCs are used for redefining distance information among images.

Formally, each CC is defined as a set of images $\mathcal{C}_l$. Given two any images img_i, $img_j \in \mathcal{C}_l$, there is a path between img_i and img_j. Algorithms for search in graphs, both Breadth First Search (BFS) and Depth First Search (DFS) can be used for computing the CCs. The output for the entire dataset is given by a set of CCs $\mathcal{S} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m\}$, such that $\{\mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_m\} = \mathcal{C}$ and $\{\mathcal{C}_1 \cap \mathcal{C}_2 \cap \cdots \cap \mathcal{C}_m\} = \emptyset$. Notice that the threshold t_k and the number of Connected Components m are related: as it grows the value of t_k, the graph becomes more connected, decreasing m.

3.4. Reciprocal kNN Graph CCs Distance

Both the edges of the Reciprocal kNN Graph G_r and the set of Connected Components $\mathcal{S}$ encode evidences of similarity among images. Therefore, such information is jointly exploited to compute a similarity score among images. The graph G_r is updated for different depths of reciprocal neighborhood. For each depth $t_k \leq k$, the similarity scores are increased, such that higher weights are assigned to neighbors at top positions (smaller t_k values).

Formally, a similarity scores $w_e(i,j)$ between images img_i, img_j is defined based on the graph connectivity. Each collection image $img_q \in \mathcal{C}$ which present edges to img_i and img_j represent an similarity increase between them. Let $E(q)$ denotes the set of nodes to which img_q has edges at a given t_k, the score is defined as:

$$w_e(i,j) = \sum_{t_k=1}^{k} \sum_{q \in \mathcal{C} \wedge i,j \in E(q)} (k - t_k + 1). \tag{5}$$

Analogously, a similarity scores $w_c(i,j)$ is defined based on information provided by the Connected Components. The score represent a similarity increase when images img_i, img_j are in the same CC. The score is also defined considering different t_k values, as follows:

$$w_c(i,j) = \sum_{tk=1}^{k} \sum_{i,j \in \mathcal{C}_l} (k - t_k + 1). \tag{6}$$

The final similarity score $w_c(i,j)$ is defined considering both edges and CCs information:

$$w(i,j) = w_e(i,j) + w_c(i,j). \tag{7}$$

Finally, a *Reciprocal kNN Graph CCs Distance* ρ_r is computed inversely proportional to the similarity score, as:

$$\rho_r(i,j) = \frac{1}{1 + w(i,j)}. \tag{8}$$

Based on the distance ρ_r a new and more effective set of ranked lists $\mathcal{T}_r$ can be computed. Notice that both the input and the output of the algorithm are defined in terms of set of ranked lists. Therefore, the algorithm can be iteratively repeated, further improving the retrieval effectiveness.

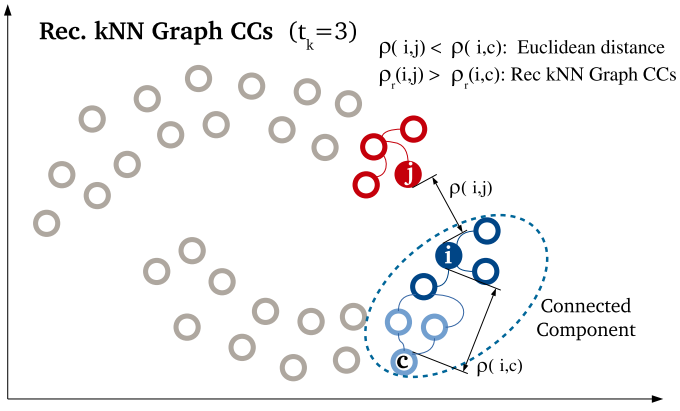


Fig. 3. Representation of the Reciprocal kNN Graph and CCs.

Let the superscript $^{(t)}$ denotes the iteration, we can define the distance of next iteration in terms of the current similarity core:

$$\rho^{(t+1)}(i,j) = \frac{1}{1 + w(i,j)^{(t)}}. \quad (9)$$

After the execution along T iterations, a final distance ρ_r can be obtained as $\rho_r(i,j) = \rho^{(T)}(i,j)$, leading to a definitive set of ranked lists.

Fig. 3 illustrates the analysis performed by the Reciprocal kNN Graph and CCs, considering the *Two Moons* dataset. The initial distance measure is defined by the Euclidean distance between the points. Three points (i, j, and c) are taken as examples, whose edges are highlighted. According to the Euclidean distance, i is nearest to j, and consequently more similar to j than c. In the proposed approach, i and c are assigned to the same connected component, what leads to an increase of similarity between them. Therefore, on the contrary of the Euclidean distance, the proposed method defines i more similar to c, respecting the dataset manifold and its geometric structure.

3.5. Reciprocal kNN Graph CCs algorithm

A simple, yet effective and efficient algorithm can be derived for computing the graph-based similarity score used by the proposed unsupervised manifold learning method.

Algorithm 1 outlines an approach for computing the similarity score among images. Different depths of neighboorhood, with $1 \leq t_k \leq k$, are considered in the loop of Lines 3–19. The similarity increments are computed considering two situations:

- Among set of images which composes the reciprocal adjacency of img_q (Lines 7–11);
- Among set of images in a connected component $\mathcal{C}_l$(Lines 13–17).

Based on the similarity score, a new distance is computed and the ranked lists are updated by a stable sorting algorithm until the L positions. Once the Reciprocal kNN Graph is very sparse, the time complexity required for computing the Connected Components and similarity scores is $O(n)$. Other steps required for the method computation, as the rank normalization or the sorting step are restricted to the top-L positions, also leading to a $O(n)$ complexity.

3.6. Rank fusion

Visual information extracted from different features encode diverse and complementary aspects about images. In last decades, diverse visual retrieval approaches have been proposed [2,5,6], including global features, mid-level representations and convolutional neural network-based. Therefore, considering different features which produce effective rankings in isolation and are complementary to each other, it is expected that a higher search accuracy can be achieved by combining them [44]. Actually, various recent retrieval methods [23,24,44] have used fusion approaches, achieving high-effective results in image retrieval tasks.

In this work, the capacity of the Reciprocal kNN Graph and CCs for discoverying the dataset manifold is also exploited for fusion tasks. Diverse features can properly discover distinct and complementary parts of the dataset manifold, enhancing the effectiveness of retrieval. For this purpose, we propose a rank fusion approach based on a linear combination of similarity scores computed for different features at the first iteration.

Let $\mathcal{D} = \{D_1, D_2, \ldots, D_d\}$ be a set of different image descriptors. Let $w_c(i,j)^{(1)}$ denotes the similarity score computed for the descriptor D_c, where the superscript $^{(1)}$ indicates the first iteration of the algorithm.

The fused similarity score $w_f(q,i)^{(1)}$ is defined as follows:

$$w_f(q,i)^{(1)} = \sum_{c=1}^{d} w_c(i,j)^{(1)}. \quad (10)$$

Based on a single and fused similarity score, a distance $\rho_r(i,j)^{(1)}$ (Eq. (9)) can be computed, giving rise to a new set of ranked lists. This distance is used for the next iterations of the manifold learning algorithm, computed in the same way as for a single descriptor.

Algorithm 1 Reciprocal kNN Graph Similarity Score.

Require: Set of Ranked lists $\{\tau_1, \tau_2, \ldots, \tau_n\}$, Parameter k
Ensure: Reciprocal kNN Graph Similarity Score $w(i,j)$

```
1:  t_k ← 1
2:  G_r ← buildReciprocalkNNGraph ()
3:  while t_k ⩽ k do
4:    G_r ← updateReciprocalkNNGraph (t_k)
5:    S ← extractConnectedComponents (G_r)
6:    { Similarity from Graph Edges }
7:    for all img_q ∈ V do
8:      for all img_i, img_j ∈ E(q) do
9:        w(i,j) ← w(i,j) + (k − t_k + 1)
10:     end for
11:   end for
12:   { Similarity from Connected Components }
13:   for all C_l ∈ S do
14:     for all img_i, img_j ∈ C_l do
15:       w(i,j) ← w(i,j) + (k − t_k + 1)
16:     end for
17:   end for
18:   t_k ← t_k + 1
19: end while
```

4. Experimental evaluation

This section discusses the experimental evaluation conducted for assessing the effectiveness of the proposed manifold learning approach. A large set of experiments were performed following a rigorous experimental protocol. The method were evaluated under diverse conditions, involving different retrieval tasks, 6 public datasets and 21 different image descriptors.

Section 4.1 describes the datasets, features and experimental protocol used. Section 4.2 discusses the impact of parameter values. Section 4.3 presents the experimental results for the proposed approach considering various shape, color, and texture descriptors. Sections 4.4 and 4.5 presents the experimental results for object

Table 1
Datasets used in the experimental evaluation.

Dataset	Size	Type	General description	Effectiv. Measure
MPEG-7 [45]	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 [46]	280	Color Scenes	Dataset composed of images from 7 soccer teams, containing 40 images per class.	MAP
Brodatz [47]	1,776	Texture	A popular dataset for texture descriptors evaluation composed of 111 different textures divided into 16 blocks.	MAP
ETH-80 [48]	3280	Objects	Dataset equally divided into 8 classes, with images containing one single object.	MAP
Holidays [49]	1491	Scenes	Commonly used as image retrieval benchmark, the dataset is composed of 1491 personal holiday pictures with 500 queries.	MAP
UKBench [50]	10,200	Objects/Scenes	Popular benchmark, composed of 2550 objects or scenes. Each object/scene is captured 4 times from different viewpoints, distances, and illumination conditions.	N-S Score

Fig. 4. Samples images from six dataset considered.

retrieval and natural image tasks, respectively. Visual results are analyzed in Section 4.6 and the main components of the method are individually evaluated in Section 4.7. Finally, experiments aiming at comparing our results to state-of-the-art related methods are presented in Section 4.8.

4.1. Datasets, features and experimental protocol

The proposed manifold learning algorithm was evaluated on six well-known public datasets, with diverse types of images, diverse characteristics and different sizes. Table 1 presents a summary of datasets used and Fig. 4 illustrates sample images from each dataset.

Several image descriptors are used, including local, global (shape, color, and texture properties), and convolutional neural network-based features. Table 2 describes the features used for each dataset. The objective is evaluate the robustness of the manifold learning algorithm for improving the effectiveness of different image features.

All images are considered as query images for most of datasets, except for Holidays [49], which uses 500 queries for comparison purposes. The Mean Average Precision (MAP) is used as effectiveness measure for most of experiments. Other evaluation measures are used for comparisons with other approaches: the N-S score [50] is used for UKBench [50] dataset and the Recall at 40 (bull's eye score) for MPEG-7 [45] dataset. The relative gains are reported for most of experiments. Let M_b, M_a be the effectiveness measure respectively before and after the use of the manifold learning algorithm, the relative gain is defined as $G = (M_a - M_b)/M_b$. Statistical paired t-tests were also conducted for assessing the statistical significance of retrieval results before and after the use of the proposed method.

Table 2
Image descriptors considered for each dataset.

Dataset	Image Features	Type
Soccer [46]	Global Color Histogram (GCH) [51], Auto Color Correlograms (ACC) [52], Border/Interior Pixel Classification (BIC) [53]	Color
MPEG-7 [45]	Segment Saliences (SS) [54], Beam Angle Statistics (BAS) [55], Inner Distance Shape Context (IDSC) [56], Contour Features Descriptor (CFD) [57], Aspect Shape Context (ASC) [58], Articulation-Invariant Representation (AIR) [59]	Shape
Brodatz [47]	Local Binary Patterns (LBP) [60], Color Co-Occurrence Matrix (CCOM) [61], Local Activity Spectrum (LAS) [62]	Texture
ETH-80 [48]	ACC [52], BIC [53], GCH [51], and Color Structure Descriptor (CSD) [63]	Color
Holidays [49]	Joint Composite Descriptor (JCD) [64], Scalable Color Descriptor (SCD) [65] Color and Edge Directivity Descriptor Spatial Pyramid (CEED-Spy) [66,67], ACC [52], Convolutional Neural Network by Caffe [3] (CNN-Caffe), Convolutional Neural Network by OverFeat [4] (CNN-OverFeat)	Color, Texture, BoVW, CNN
UKBench [50]	CEED-Spy [66,67], Fuzzy Color and Texture Histogram Spatial Pyramid (FCTH-SPy) [67,68], SCD [65], ACC Spatial Pyramid (ACC-SPy) [52,67], CNN-Caffe [3] ACC [52], Vocabulary Tree (VOC) [69]	Color, Texture, BoVW, CNN

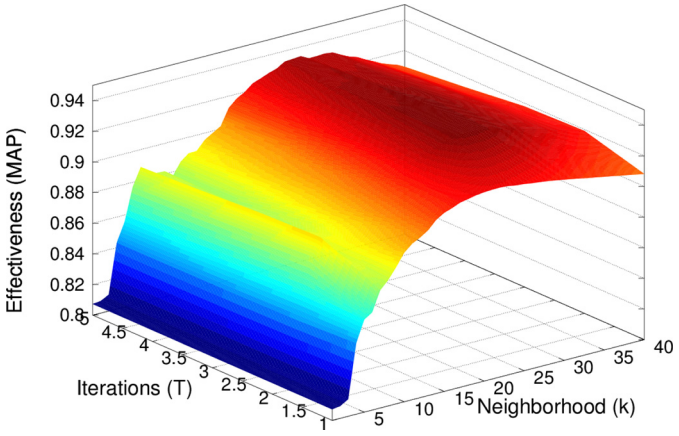


Fig. 5. Impact of neighborhood size and iterations on effectiveness.

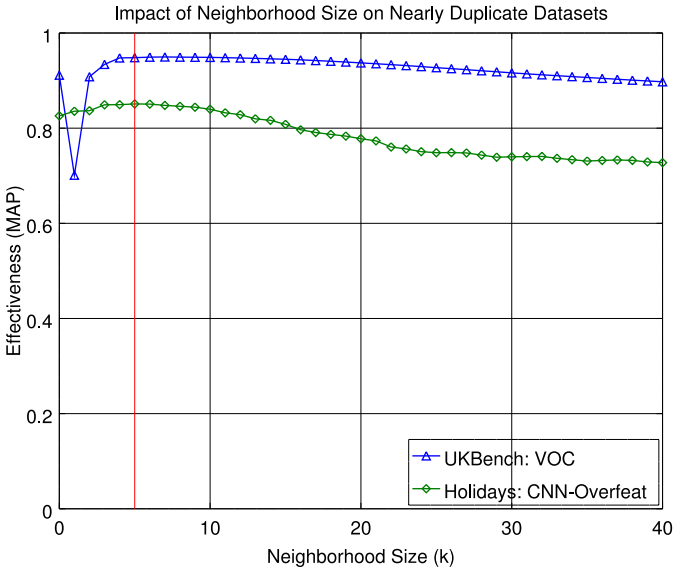


Fig. 6. Impact of neighborhood size on nearly duplicate datasets.

4.2. Impact of parameters

The proposed approach consider only two parameters: k, which defines the size of the neighborhood and T, which indicates the number of iterations. The algorithm also consider the value L, which determines the size of ranked list and a trade-off between effectiveness and efficiency. In all experiments, the value of L is defined in terms of k, as $L = 4 \times k$.

An experiment was conducted aiming at analyzing the impact of k and T on effectiveness results. The MPEG-7 [45] and the CFD [57] shape descriptor were considered for the experiment. The MAP is used as effectiveness measure, evaluated in function of k and T.

Fig. 5 illustrates the MAP results according to different values of k and T. It can be observed a small variation for different iterations, indicating that the most expressive effectiveness gains are obtained at the first iteration. Additionally, a large red region demonstrates the robustness of the method to different parameters settings.

For most of experiments, we report the effectiveness scores in two scenarios: using fixed parameters values ($k = 20$ and $T = 1$)[1] and using the best parameter combination in predefined intervals (k in [5,40] and T in [1,5]). The objective is to demonstrate the potential of the method for obtaining high effectiveness gains and, at same time, evaluate the method in adverse situations, when there is no information about the retrieval task.

[1] For UKBench and Holidays datasets, parameters are defined as $k = 5$ and $T = 2$ due to the small number of images per class.

An experiment was also conducted in order to evaluate the effectiveness of the proposed method according to different values of k on all the six datasets. The most effective descriptor on each dataset was used in the experiment and the parameter k was varied in the interval [0,40]. The results are organized according to the type of dataset: nearly duplicate and general datasets, as also suggested by other works [23]. Figs. 6 and 7 illustrates the results for nearly duplicate and general datasets, respectively. The vertical red lines indicate the parameter values used in the experiments.

In general, a similar behavior can be observed on most of datasets. The effectiveness results presents an instability at the beginning of the curve, which subsequently reaches an stabilization. On nearly duplicate datasets the stabilization occurs for lower values (around $k = 5$) and on general datasets for higher values of k (around $k = 20$).

4.3. Shape, color, and texture retrieval

The effectiveness of the proposed manifold learning algorithm is first evaluated in general image retrieval tasks, considering shape, color, and texture features. The results are described as following:

- **Shape retrieval:** The experiments for shape retrieval considering the MPEG-7 [45] dataset and six different descriptors are

Table 3
Reciprocal kNN Graph Distance and CCs for various image retrieval tasks. Mean Average Precision (MAP) considering shape, color, and texture features.

Descriptor	Dataset	Original **score (MAP)**	Rec. kNN **Graph CCs** k = 20; T = 1	Rec. kNN **Graph CCs** Best params	Best **params:** k; T	Relative **gain**	Stat. **sig. 99%**
Shape descriptors							
SS [54]	MPEG-7	37.67%	47.84%	53.02%	30; 3	+40.75%	•
BAS [55]	MPEG-7	71.52%	82.30%	84.08%	25; 3	+17.56%	•
IDSC [56]	MPEG-7	81.70%	89.65%	91.26%	23; 3	+11.70%	•
CFD [57]	MPEG-7	80.71%	92.33%	93.83%	25; 2	+16.26%	•
ASC [58]	MPEG-7	85.28%	92.62%	93.57%	23; 3	+9.72%	•
AIR [59]	MPEG-7	89.39%	97.17%	97.81%	37; 2	+9.42%	•
Color descriptors							
GCH [51]	Soccer	32.24%	33.53%	34.90%	40; 2	+8.25%	•
ACC [52]	Soccer	37.23%	40.91%	44.94%	40; 2	+20.71%	•
BIC [53]	Soccer	39.26%	42.85%	45.99%	40; 2	+17.14%	•
Texture descriptors							
LBP [60]	Brodatz	48.40%	50.82%	51.77%	17; 2	+6.96%	•
CCOM [61]	Brodatz	57.57%	64.27%	66.68%	21; 3	+15.82%	•
LAS [62]	Brodatz	75.15%	80.57%	81.97%	18; 3	+9.08%	•

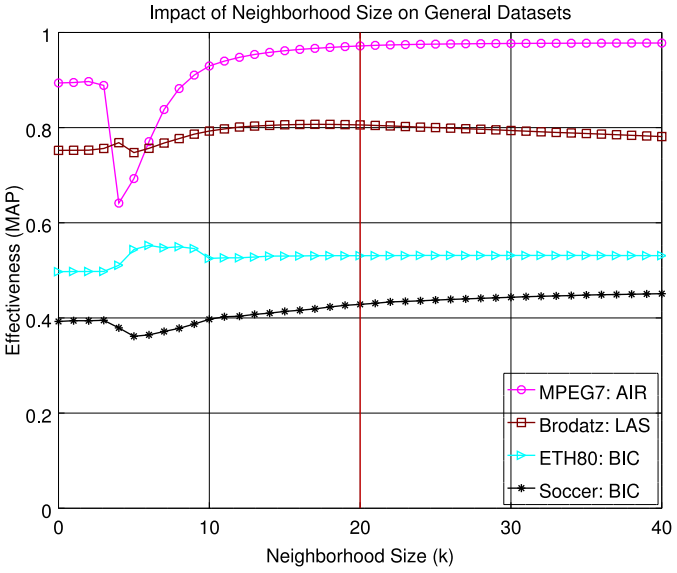


Fig. 7. Impact of neighborhood size on general datasets.

Table 4
Reciprocal kNN Graph Distance + CCs for rank aggregation tasks considering shape, color, and texture retrieval tasks.

Descriptor	Type	Dataset	Parameters	Score **(MAP)**
CFD [57]	Shape	MPEG-7	–	80.71%
ASC [58]	Shape	MPEG-7	–	85.28%
AIR [59]	Shape	MPEG-7	–	89.39%
CFD+ASC	Shape	MPEG-7	$k = 20$; $T = 1$	97.68%
CFD+ASC	Shape	MPEG-7	$k = 25$; $T = 2$	99.18%
CFD+AIR	Shape	MPEG-7	$k = 20$; $T = 1$	99.32%
CFD+AIR	Shape	MPEG-7	$k = 37$; $T = 2$	**100%**
AIR+ASC	Shape	MPEG-7	$k = 20$; $T = 1$	99.16%
AIR+ASC	Shape	MPEG-7	$k = 37$; $T = 2$	99.89%
CFD+ASC+AIR	Shape	MPEG-7	$k = 20$; $T = 1$	99.79%
CFD+ASC+AIR	Shape	MPEG-7	$k = 37$; $T = 2$	**100%**
ACC [52]	Color	Soccer	–	37.23%
BIC [53]	Color	Soccer	–	39.26%
BIC+ACC	Color	Soccer	$k = 20$; $T = 1$	43.19%
BIC+ACC	Color	Soccer	$k = 40$; $T = 2$	**46.05%**
CCOM [61]	Texture	Brodatz	–	57.57%
LAS [62]	Texture	Brodatz	–	75.15%
LAS+CCOM	Texture	Brodatz	$k = 20$; $T = 1$	81.49%
LAS+CCOM	Texture	Brodatz	$k = 18$; $T = 3$	**84.72%**

presented in Table 3. Significant positive gains can observed for all descriptors, ranging from +9.42% to **+40.75%**. For example, the effectiveness of SS [54] shape descriptor was improved from 37.67% to 53.02%.[2]

- **Color retrieval:** The experiments involving color retrieval were conducted on Soccer dataset [46], considering three different descriptors. Table 3 presents the experimental results, where positive gains were obtained for all color descriptors, ranging from +8.25% to +20.71%.
- **Texture retrieval:** Table 3 also presents the effectiveness results obtained for three different texture descriptors on the Brodatz [47] dataset. We can observe positive gains ranging from +6.96% to +15.82%. The obtained gains are statistical significant for all considered descriptors.

Notice that effectiveness scores obtained by fixed parameters values and the best parameters combination are very similar, demonstrating the robustness of the proposed approach to parameters settings.

[2] The relative gains reported refer to the results obtained by the best parameters.

The proposed manifold learning algorithm was also evaluated for rank fusion tasks, on shape, color, and texture retrieval. Two visual descriptors are considered for color and texture, and three for shape retrieval, selected according to the best effectiveness results in isolation.

Table 4 presents the results for rank fusion tasks, considering fixed parameter values ($k = 20$ and $T = 1$) and the best combination of parameters of the best descriptor. The original MAP results obtained by each descriptor in isolation are also reported for comparison purposes. As it can be observed, very high effective results are obtained. Considering texture retrieval, the fusion of descriptors with MAP scores of 57.57% and 75.15% leads to 84.72%. For shape retrieval, a MAP score of 100% is obtained, which indicates perfect retrieval results.

4.4. Object retrieval

Experiments conducted for object retrieval tasks considered four color descriptors on the ETH-80 [48] dataset. All experiments considered fixed values of parameters ($k = 20$ and $T = 1$). Table 5 presents the MAP scores of each descriptor.

Table 5
Reciprocal kNN Graph Distance + CCs on ETH-80 [48] dataset.

Descriptor	Score (MAP)	Rec kNN Graph CCs	Relative gain	Stat. sig. 99%
BIC [53]	49.72%	53.08%	+6.76%	•
ACC [52]	48.50%	52.22%	+7.67%	•
CSD [63]	48.46%	52.08%	+7.47%	•
GCH [51]	41.62%	43.87%	+5.41%	•

Table 6
Reciprocal kNN Graph Distance + CCs on UKBench [50] dataset.

Descriptor	N-S Score	Rec kNN Graph CCs	Relative gain
JCD [64]	2.79	2.97	+6.45%
CEED-SPy [66,67]	2.81	3.09	+9.96%
FCTH-SPy [67,68]	2.91	3.18	+9.27%
SCD [65]	3.15	3.36	+6.67%
CNN-Caffe [3]	3.31	3.60	+8.76%
ACC [52]	3.36	3.59	+6.85%
VOC [69]	3.54	3.76	+6.21%
VOC+ACC	–	3.88	+9.60%
VOC+CNN-Caffe	–	3.89	+9.89%
ACC+CNN-Caffe	–	3.85	+8.76%
VOC+ACC+CNN-Caffe	–	**3.93**	+11.02%

Positive gains were obtained for, ranging from +5.41% to +7.67%. Statistical significance can be observed for all descriptors. The MAP scores are improved, for example, from 48.50% to 52.22%.

4.5. Natural image retrieval

The proposed method was also evaluated in natural image retrieval tasks, considering two popular datasets, commonly used as benchmark in image retrieval tasks: the University of Kentucky Recognition Benchmark - UKBench [50] and the Holidays [49] dataset.

The UKBench dataset is composed of 2550 objects or scenes captured 4 times from different viewpoints, totalling 10,200 images. The N-S score is used as evaluation measure, computed between 1 and 4, which corresponds to the number of relevant images among the first four image returned. Due to the small number of images per class, the UKBench [50] dataset is very challenger for unsupervised manifold learning and post-processing methods.

Different descriptors were considered, including various color and color/texture global descriptors, extracted by the LIRE framework [67]. Local descriptors are also exploited, using the rank positions[3] obtained by a vocabulary tree based retrieval (VOC) [50,69], which uses SIFT features. Convolution neural networks features are extracted from the 7th layer using the Caffe framework [3]. A 4096-dimensional CNN-Caffe descriptor was considered for each input image resized of 256 × 256 pixels, and the Euclidean distance was used for computing the rankings.

Table 6 presents the results obtained by the proposed method on UKBench [50] dataset. Significant positive gains are obtained, reaching +9.96%. For example, the N-S score obatined by CNN-Caffe [3] feature was improved from 3.31 to 3.60. Even more remarkable results can be observed for fusion tasks, reaching a N-S score of **3.93**. The relative gains are computed in comparison with the best descriptor in isolation.

The experiments conducted on the Holidays [49] dataset considered analogous conditions. Instead of local descriptors, we used other CNN feature: Overfeat [4]. Once the number of relevant images per class is smaller, unsupervised manifold learning tasks are even challenger. Table 7 presents the results on the Holidays [49] dataset. Effectiveness gains can be observed for all features, reaching a MAP score of **86.19%** of the fusion of ACC color and CNN-Overfeat features.

[3] http://research.rutgers.edu/~shaoting/image_search.html (As of September 2016).

Table 7
Reciprocal kNN Graph Distance + CCs on the Holidays [49] dataset.

Descriptor	Original MAP	Rec kNN Graph CCs	Relative gain
JCD [64]	52.83%	53.69%	+1.63%
SCD [65]	54.26%	56.50%	+4.13%
FCTH-SPy [67,68]	55.42%	58.92%	+6.32%
CEED-SPy [66,67]	56.09%	59.02%	+5.22%
CNN-Caffe [3]	64.09%	68.47%	+6.83%
ACC [52]	64.29%	68.80%	+7.02%
CNN-OverFeat [4]	82.59%	84.99%	+2.91%
ACC + CEED-SPy	–	71.90%	+11.84%
ACC + CNN-Caffe	–	78.93%	+22.77%
ACC + CNN-Caffe + CNN-Overfeat	–	85.73%	+3.80%
ACC + CNN-OverFeat	–	**86.19%**	+4.35%

4.6. Qualitative and visual analysis

Once the quantitative results demonstrated significant gains for effectiveness results, this section conduct a qualitative analysis using visual representations for evaluating the impact of the use of the proposed manifold learning algorithm.

First, we analyzed the impact of the proposed manifold algorithm on distance distribution using a bidimensional representation of a dataset before and after the execution of the algorithm. For the representation, two arbitrary images are selected, named as reference images. Next, all collection images are represented in the bidimensional space, such that their position is defined according to their distance to the reference images. Formally, given two reference images img_i and img_j and an image img_l that is represented in the bidimensional space, the position (x, y) of img_l is defined as $(\rho(img_i, img_l), \rho(img_j, img_l))$.

Fig. 8 (a) illustrates the reference images (*"tree-13"*, *"tree-7"*). Fig. 8 (b) illustrates the distance distribution obtained by the CFD [57] shape descriptor from the MPEG-7 [45] dataset. Similar images to the reference images are illustrated in red circles and remaining images in blue. As we can observe, similar and non-similar images (red circles and blue crosses) are mixed in the distance space. Fig. 8 (c) illustrates the distance after the execution of the proposed manifold learning algorithm. Notice the capacity of the proposed algorithm of considering the dataset manifold, which increases the separability between similar and non-similar images.

The visual effect of the new distance distribution on image retrieval tasks is showed in Fig. 9. Visual retrieval results obtained by CFD [57] shape descriptor on MPEG-7 [45] dataset before and after the use of the algorithm are illustrated. Three queries are considered (*tree-13*, *tree-7* and *turtle-29*) and are represented with green borders. Incorrect retrieval results are illustrated with red borders. The precision of retrieval at top-20 positions was improved from values between 15% and 20% to 100% after the use of the manifold learning algorithm.

The impact of the dataset manifold in the effectiveness of image retrieval tasks is illustrated in Figs. 10 and 11. Each point represent one image of MPEG-7 [45] dataset. The x-coordinate represents the initial average precision obtained by SS [54] and CFD [57] shape descriptors. The y-coordinate is defined by the average precision after the use of the proposed manifold learning algorithm. As can be observed, the vast majority of points are located above the main diagonal, representing the effectiveness gains in retrieval results.

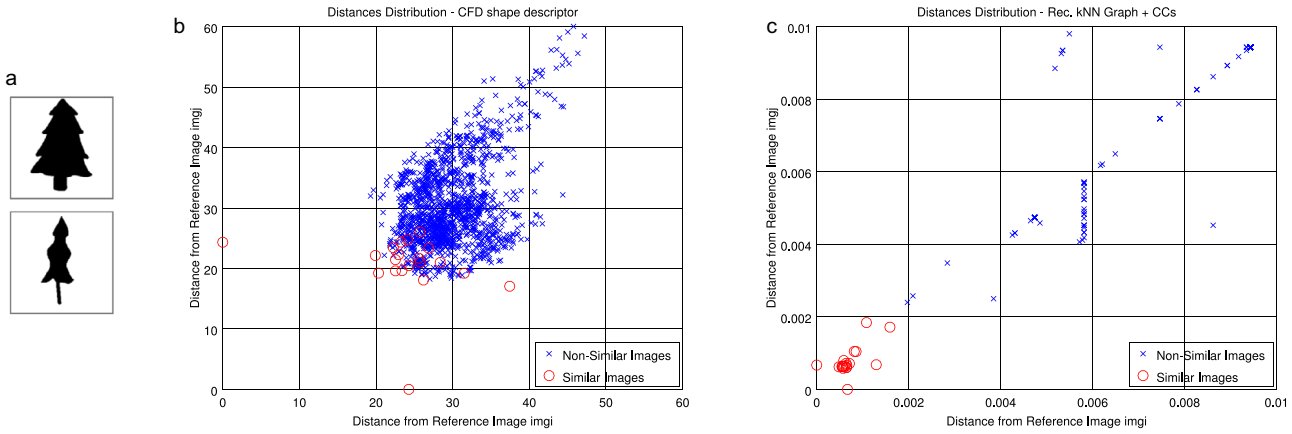


Fig. 8. Impact of the algorithm on distances distribution for similar reference images: **(a)** Similar reference Images img_i and img_j (*tree-13.gif* and *tree-7.gif*) from the MPEG-7 [45] dataset; **(b)** Original distances distribution given by CFD [57] shape descriptor; **(c)** Distances distribution defined by the Rec. kNN Graph + CCs algorithm.

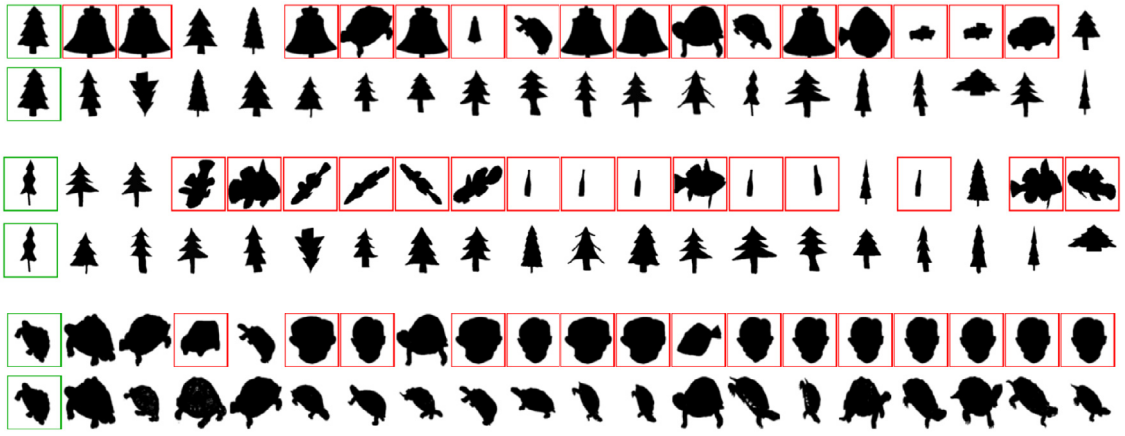


Fig. 9. Retrieval results before and after the algorithm for two different queries using the CFD [57] shape descriptor. Query images are represented with green borders and wrong images with red borders. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

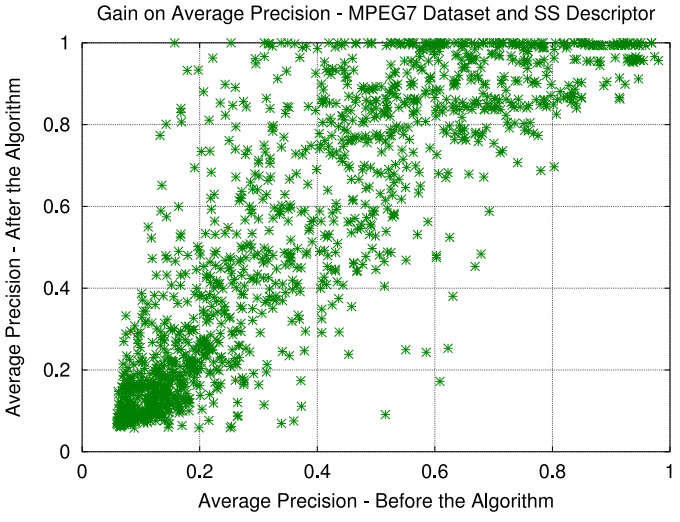


Fig. 10. Gains of precision: SS [54].

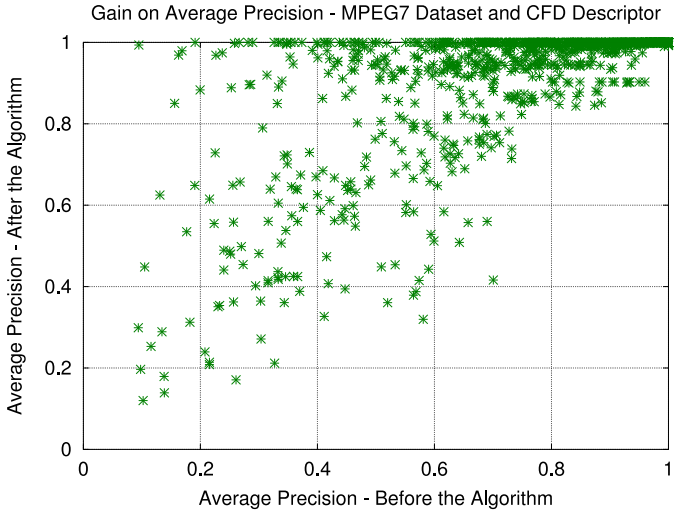


Fig. 11. Gains of precision: CFD [57].

The benefits of the use of the algorithm is also analyzed on the UKBench [50] dataset, considering different features. Fig. 12 illustrates the retrieval results obtained by ACC [52], VOC [69], CNN-Caffe [3], and by the proposed manifold learning algorithm. Relevant images are represented with green borders. We can observe that, even based on low-effective results given by the features, the algorithm produces perfect top-4 rankings.

4.7. Analysis of algorithm components

An experiment was conducted to evaluate the main components of the proposed method in isolation, aiming at measuring

Fig. 12. Visual retrieval examples on UKBench [50] dataset.

their individual contributions to the effectiveness gains. The experiment was conducted on the six datasets, considering the most effective descriptor and the descriptor which presented the higher effectiveness gain on each dataset. Three different scenarios are considered: (*i*) only similarity information provided by the edges of the Rec. kNN Graph; (*ii*) only information obtained from Connected Components; and (*iii*) the full algorithm, considering both information, including rank normalization. Table 8 presents the experimental results, with the best results for each descriptor highlighted in bold. The full algorithm yielded the best scores for most of descriptors, outperformed by CCs on only one dataset. However, in opposite to the edges results, the CCs results are very unstable to different descriptors-datasets. Such behavior and the higher average results obtained by the full algorithm indicate the complementarity of the Rec. kNN Graph and CCs information.

Table 8
Impact of different algorithm components on effectiveness.

Descriptor	Dataset	Original MAP	Only edges	Only CCs	Full algorithm
CFD [57]	MPEG-7	80.71%	90.06%	88.73%	**92.33%**
AIR [59]	MPEG-7	89.39%	96.94%	55.84%	**97.17%**
ACC [52]	Soccer	37.23%	40.54%	33.43%	**40.91%**
BIC [53]	Soccer	39.26%	42.09%	32.97%	**42.85%**
CCOM [61]	Brodatz	57.57%	64.05%	50.83%	**64.27%**
LAS [62]	Brodatz	75.22%	79.71%	76.34%	**80.57%**
ACC [52]	ETH-80	48.52%	49.03%	**52.42%**	52.22%
BIC [53]	ETH-80	49.76%	50.27%	**53.83%**	53.08%
ACC [52]	Holidays	64.29%	66.60%	66.85%	**69.47%**
CNN-OverFeat [4]	Holidays	82.59%	84.69%	84.50%	**85.09%**
FCTH-SPy [67,68]	UKBench	77.81%	80.53%	79.07%	**81.82%**
VOC [69]	UKBench	91.14%	94.03%	93.63%	**94.82%**
Average		66.12%	69.88%	64.04%	**71.22%**

Table 9
Comparison with post-processing methods on the **MPEG-7** [45] dataset.

Shape descriptors	(Bull's eye score)	
CFD [57]	–	84.43%
IDSC [56]	–	85.40%
ASC [58]	–	88.39%
AIR [59]	–	93.67%
Post-processing methods		
Algorithm	Descriptor(s)	Bull's eye score
Contextual Dissimilarity Measure [42]	IDSC [56]	88.30%
Graph Transduction [70]	IDSC [56]	91.00%
Self-Smoothing Operator [22]	IDSC [56]	92.77%
Local Constr. Diff. Process [26]	IDSC [56]	93.32%
Shortest Path Propagation [27]	IDSC [56]	93.35%
Mutual kNN Graph [71]	IDSC [56]	93.40%
SCA [19]	IDSC [56]	93.44%
Rec. kNN Graph CCs	**IDSC** [56]	**93.62%**
Index-Based Re-Ranking [72]	CFD [57]	92.85%
Correlation Graph [43]	CFD [57]	94.27%
RL-Sim [34]	CFD [57]	94.27%
Rec. kNN Graph CCs	**CFD**[57]	**96.51%**
Generic Diffusion Process [21]	ASC [58]	93.95%
Index-Based Re-Ranking [72]	ASC [58]	94.09%
Correlation Graph [43]	ASC [58]	95.22%
Local Constr. Diff. Process [26]	ASC [58]	95.96%
Rec. kNN Graph CCs	**ASC**[58]	**96.04%**
Tensor Product Graph [20]	ASC [58]	96.47%
RL-Sim [34]	AIR [59]	99.94%
Tensor Product Graph [20]	AIR [59]	99.99%
Generic Diffusion Process [21]	AIR [59]	100%
Neighbor Set Similarity [73]	AIR [59]	100%
Rec. kNN Graph CCs	**AIR**[59]	**100%**

4.8. Comparison with other approaches

The proposed method is also evaluated in comparison with various state-of-the-art unsupervised learning methods and recently proposed retrieval approaches. Experiments were conducted on three image datasets: MPEG-7 [45], Holidays [49] and UKBench [50], which are popular datasets commonly used as benchmark for image retrieval and post-processing methods.

Table 9 presents the results on the MPEG-7 [45] dataset in comparison with several state-of-the-art post-processing methods. The

Table 10
Comparison with state-of-the-art on the **Holidays** [49] dataset.

MAP scores for recent retrieval methods.					
Jégou et al. [49]	Tolias et al. [74]	Qin et al. [75]	Zheng et al. [76]	Zheng et al. [77]	Li et al. [78]
75.07%	82.20%	84.40%	85.20%	85.80%	89.20%

MAP scores for the proposed method		
Descriptor	Baseline: *Graph Fusion*[23]	Proposed: *Rec. kNN Graph CCs*
ACC	66.42%	68.80%
CNN-Caffe	66.79%	68.47%
CNN-Overfeat	83.79%	84.99%
ACC+CNN-Caffe	71.02%	78.93%
ACC+CNN-Overfeat	76.55%	**86.19%**
ACC+CNN-Caffe+CNN-Overfeat	80.06%	85.73%

Table 11
Comparison with state-of-the-art on the **UKBench** [50] dataset.

N-S scores for recent retrieval methods						
Zheng et al. [79]	Qin et al. [38]	Wang et al. [80]	Zhang et al. [23]	Zheng et al. [44]	Bai et al. [19]	Xie et al. [81]
3.57	3.67	3.68	3.83	3.84	3.86	3.89

N-S scores for the proposed method		
Descriptor	Baseline: *Graph Fusion*[23]	Proposed: *Rec. kNN Graph CCs*
ACC	3.48	3.59
CNN	3.45	3.60
VOC	3.67	3.76
ACC+CNN	3.70	3.88
ACC+VOC	3.78	3.88
VOC+CNN	3.78	3.89
ACC+VOC+CNN	3.86	**3.93**

MPEG-7 dataset is very frequently used as benchmark for unsupervised post-processing methods in the literature. The bull's eye score (Recall@40), which counts all matching shapes within the top-40 ranked images, is used as evaluation measure. The proposed approach is evaluated in comparison with diverse post-processing methods considering the same features, given by four shape descriptors: IDSC [56], ASC [58], CFD [57], and AIR [59]. We can observe that the proposed algorithm achieves the best scores for most of them, including a bull's eye score of **100%** achieved for the AIR [59] descriptor.

Table 10 shows the MAP scores obtained on the Holidays [49] dataset, in comparison with various recent image retrieval approaches. Once different features are used by the compared approaches, a comparison considering the same features used by our method is also included. The Graph Fusion [23], which is a recent and relevant unsupervised related method, is used as baseline. The parameters settings was followed the same values used by the proposed method ($k = 5$) and fusion tasks was perfomed by the graph density variation [23]. As we can observe, the effectiveness results are very high, superior to the most of considered methods.

Table 11 presents the results obtained by the proposed manifold learning algorithm results on the UKBench [50] dataset. The UKBench [50] dataset is often used as benchmark for both general retrieval approaches and unsupervised post-processing methods. Analogous to the Holidays dataset, a comparison with the Graph Fusion [23] method using the same features is also presented. The proposed Rec. kNN Graph + CCs algorithm yielded a N-S scores of **3.93**, the best score in comparison with other recent state-of-the-art methods.

5. Conclusions

A novel unsupervised manifold learning algorithm is proposed in this paper, employing a graph-based approach to consider the geometry of the dataset in order to learn a new distance. The proposed Reciprocal kNN Graph exploits the Connected Components to analyze the information encoded in the reciprocal ranking references. In this way, a simple but effective algorithm is derived for recomputing the distance among images taking into account the dataset manifold.

A broad experimental evaluation was conducted, involving various experiments, different retrieval tasks and several datasets and image features. Experimental results demonstrated the robustness of the method, once significant effectiveness gains were achieved in diverse retrieval scenarios, considering 6 different datasets and 21 image features. Very high retrieval performance was reached in various datasets traditionally used as benchmark. The method was also evaluated in comparison with state-of-the-art methods, achieving superior effectiveness results to most of considered approaches.

Future work focuses on the investigation of the use of the proposed approach for improving the accuracy of classification tasks.

Acknowledgments

The authors are grateful to FAPESP - São Paulo Research Foundation (grant #2013/08645-0) and CAPES - Coordination for Higher Education Staff Development.

References

[1] M.S. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, ACM Trans. Multimedia Comput.Commun. Appl. (TOMM) 2 (1) (2006) 1–19.
[2] O.A.B. Penatti, E. Valle, R.d. S. Torres, Comparative study of global color and texture descriptors for web image retrieval, J. Vis. Commun. Image Represent. 23 (2) (2012) 359–380.
[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R.B. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional Architecture for Fast Feature Embedding, ACM international conference on Multimedia (MM'14) (2014) 675–678.
[4] 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), (2014) pp. 512–519.
[5] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput. Surv. 40 (2) (2008) 5:1–5:60.
[6] Y. Uchida, Local feature detectors, descriptors, and image representations: a survey, CoRR abs/1607.08368 (2016) http://arxiv.org/abs/1607.08368.
[7] A. Bellet, A. Habrard, M. Sebban, A survey on metric learning for feature vectors and structured data, CoRR abs/1306.6709 (2013) https://arxiv.org/abs/1306.6709v4.
[8] J.-E. Lee, R. Jin, A.K. Jain, Rank-based distance metric learning: an application to image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2008), 2008, pp. 1–8.
[9] Y. Liu, D. Zhang, G. Lu, W.-Y. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recognit. 40 (1) (2007) 262–282.
[10] B. Thomee, M. Lew, Interactive search in image retrieval: a survey, Int. J. Multimedia Inf. Retrieval 1 (2) (2012) 71–86.
[11] 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 Recognit. Lett. 32 (1) (2011) 27–37.
[12] A.T. da Silva, J.A. dos Santos, A.X. Falcão, R. da S. Torres, L.P. Magalhães, Incorporating multiple distance spaces in optimum-path forest classification to improve feedback-based learning, Comput. Vis. Image Understanding 116 (4) (2012) 510–523.
[13] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach.Intell. 34 (4) (2012) 723–742.
[14] D.C. Xiaofei He, J. Han., Learning a maximum margin subspace for image retrieval, IEEE Trans. Knowl. Data Eng. 20 (2) (2008).
[15] V.E. Liong, J. Lu, Y. Ge, Regularized local metric learning for person re-identification, Pattern Recognit. Lett. 68, Part 2 (2015) 288–296.
[16] W. Li, Y. Wu, J. Li, Re-identification by neighborhood structure metric learning, Pattern Recognit. 61 (2017) 327–338.
[17] B. Wang, F. Pan, K.-M. Hu, J.-C. Paul, Manifold-ranking based retrieval using k-regular nearest neighbor graph, Pattern Recognit. 45 (4) (2012) 1569–1577.
[18] S.C. Hoi, W. Liu, S.-F. Chang, Semi-supervised distance metric learning for collaborative image retrieval and clustering, ACM Trans. Multimedia Comput. Commun. Appl. 6 (3) (2010) 18:1–18:26.

[19] S. Bai, X. Bai, Sparse contextual activation for efficient visual re-ranking, IEEE Trans. Image Process. (TIP) 25 (3) (2016) 1056–1069.
[20] X. Yang, L. Prasad, L. Latecki, Affinity learning with diffusion on tensor product graph, IEEE Trans. Pattern Anal. Mach.Intell. 35 (1) (2013) 28–38.
[21] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, in: IEEE Conferece on Computer Vision and Pattern Recognition (CVPR'2013), 2013, pp. 1320–1327.
[22] J. Jiang, B. Wang, Z. Tu, Unsupervised metric learning by self-smoothing operator, in: International Conference on Computer Vision (ICCV'2011), 2011, pp. 794–801.
[23] S. Zhang, M. Yang, T. Cour, K. Yu, D. Metaxas, Query specific rank fusion for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 37 (4) (2015) 803–815.
[24] D.C.G. Pedronette, O.A. Penatti, R. da S. Torres, Unsupervised manifold learning using reciprocal kNN graphs in image re-ranking and rank aggregation tasks, Image Vis. Comput. 32 (2) (2014) 120–130.
[25] D.C.G. Pedronette, R. da Silva Torres, A correlation graph approach for unsupervised manifold learning in image retrieval tasks, Neurocomputing 208 (2016) 66–79.
[26] 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, 2009, pp. 357–364.
[27] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, N. Tang, Learning context-sensitive similarity by shortest path propagation, Pattern Recognit. 44 (10–11) (2011) 2367–2374.
[28] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu, J. Luo, Efficient manifold ranking for image retrieval, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, in: SIGIR '11, 2011, pp. 525–534.
[29] I. Theodorakopoulos, G. Economou, S. Fotopoulos, C. Theoharatos, Local manifold distance based on neighborhood graph reordering, Pattern Recognit. 53 (2016) 195–211.
[30] X. Xing, Y. Yu, H. Jiang, S. Du, A multi-manifold semi-supervised gaussian mixture model for pattern classification, Pattern Recognit. Lett. 34 (16) (2013) 2118–2125.
[31] Y. Pei, F. Huang, F. Shi, H. Zha, Unsupervised image matching based on manifold alignment, IEEE Trans. Pattern Anal. Mach. Intell. 34 (8) (2012) 1658–1664.
[32] J. Lu, Y.P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[33] Y. Chen, X. Li, A. Dick, R. Hill, Ranking consistency for image matching and object retrieval, Pattern Recognit. 47 (3) (2014) 1349–1360.
[34] D.C.G. Pedronette, R. da S. Torres, Image re-ranking and rank aggregation based on similarity of ranked lists, Pattern Recognit. 46 (8) (2013) 2350–2360.
[35] L. Zheng, S. Wang, J. Wang, Q. Tian, Accurate image search with multi-scale contextual evidences, Int. J. Comput. Vis. 120 (1) (2016a) 1–13.
[36] L. Zheng, Y. Yang, Q. Tian, SIFT Meets CNN: a decade survey of instance retrieval, CoRR abs/1608.01807 (2016b) https://arxiv.org/abs/1608.01807.
[37] F. Radenovic, G. Tolias, O. Chum, CNN image retrieval learns from bow: unsupervised fine-tuning with hard examples, in: European Conference Computer Vision - ECCV 2016, 2016, pp. 3–20.
[38] D. Qin, S. Gammeter, L. Bossard, T. Quack, L. van Gool, Hello neighbor: accurate object retrieval with k-reciprocal nearest neighbors, in: CVPR'2011, 2011, pp. 777–784.
[39] D.C.G. Pedronette, O.A.B. Penatti, R.T. Calumby, R. da S. Torres, Unsupervised distance learning by reciprocal kNNdistance for image retrieval, in: International Conference on Multimedia Retrieval (ICMR'14), 2014.
[40] A. Delvinioti, H. Jgou, L. Amsaleg, M.E. Houle, Image retrieval with reciprocal and shared nearest neighbors, in: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on, 2, 2014, pp. 321–328.
[41] 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.
[42] H. Jegou, C. Schmid, H. Harzallah, J. Verbeek, Accurate image search using the contextual dissimilarity measure, IEEE Trans. Pattern Anal. Mach. Intell. 32 (1) (2010) 2–11.
[43] D.C.G. Pedronette, R. da S. Torres, Unsupervised manifold learning by correlation graph and strongly connected components for image retrieval, in: International Conference on Image Processing (ICIP'2014), 2014.
[44] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, Q. Tian, Query-adaptive late fusion for image search and person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[45] L. J. Latecki, R. Lakamper and T. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000, pp. 424–429.
[46] J. van de Weijer, C. Schmid, Coloring local feature extraction, in: European Conference on Computer Vision (ECCV'2006), Part II, 2006, pp. 334–348.
[47] P. Brodatz, Textures: A Photographic Album for Artists and Designers, Dover, 1966.
[48] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, CVPR, 2, 2003. II–409–15 vol.2
[49] H. Jegou, M. Douze, C. Schmid, Hamming embedding and weak geometric consistency for large scale image search, in: European Conference on Computer Vision, in: ECCV '08, 2008, pp. 304–317.
[50] D. Nistér, H. Stewénius, Scalable recognition with a vocabulary tree, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2006), 2, 2006, pp. 2161–2168.
[51] M.J. Swain, D.H. Ballard, Color indexing, Int. J. Comput. Vis. 7 (1) (1991) 11–32.
[52] J. Huang, S.R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlograms, in: CVPR'97, 1997, pp. 762–768.
[53] 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, 2002, pp. 102–109.
[54] R. da S. Torres, A.X. Falcão, Contour salience descriptors for effective image retrieval and analysis, Image Vis. Comput. 25 (1) (2007) 3–13.
[55] N. Arica, F.T.Y. Vural, BAS: a perceptual shape descriptor based on the beam angle statistics, Pattern Recognit. Lett. 24 (9–10) (2003) 1627–1639.
[56] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 286–299.
[57] D.C.G. Pedronette, R. da S. Torres, Shape retrieval using contour features and distance optmization, in: VISAPP'2010, 1, 2010, pp. 197–202.
[58] H. Ling, X. Yang, L.J. Latecki, Balancing deformability and discriminability for shape matching, in: ECCV'2010, 3, 2010, pp. 411–424.
[59] R. Gopalan, P. Turaga, R. Chellappa, Articulation-invariant representation of non-planar shapes, in: 11th European Conference on Computer Vision (ECCV'2010), 3, 2010, pp. 286–299.
[60] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[61] V. Kovalev, S. Volmer, Color co-occurence descriptors for querying-by-example, in: International Conference on Multimedia Modeling, 1998, p. 32.
[62] B. Tao, B.W. Dickinson, Texture recognition and image retrieval using gradient indexing, J. Vis. Comun. Image Represent. 11 (3) (2000) 327–342.
[63] B. Manjunath, J.-R. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Syst. Video Technol. 11 (6) (2001) 703–715.
[64] K. Zagoris, S. Chatzichristofis, N. Papamarkos, Y. Boutalis, Automatic image annotation and retrieval using the joint composite descriptor, in: 14th Panhellenic Conference on Informatics (PCI), 2010, pp. 143–147.
[65] B. Manjunath, J.-R. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Syst. Video Technol. 11 (6) (2001) 703–715.
[66] S.A. Chatzichristofis, Y.S. Boutalis, Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval, in: Proceedings of the 6th International Conference on Computer Vision Systems, in: ICVS'08, 2008, pp. 312–322.
[67] M. Lux, Content based image retrieval with LIRe, in: Proceedings of the 19th ACM International Conference on Multimedia, in: MM '11, 2011.
[68] S.A. Chatzichristofis, Y.S. Boutalis, FCTH: fuzzy color and texture histogram - a low level feature for accurate image retrieval, in: Ninth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '08), 2008, pp. 191–196.
[69] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, T. Han, Contextual weighting for vocabulary tree based image retrieval, in: IEEE International Conference on Computer Vision (ICCV'2011), 2011, pp. 209–216.
[70] X. Yang, X. Bai, L.J. Latecki, Z. Tu, Improving shape retrieval by learning graph transduction, in: European Conference on Computer Vision (ECCV'2008), 4, 2008, pp. 788–801.
[71] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: Asian Conference on Computer Vision (ACCV'2009), 2009, pp. 655–666.
[72] D.C.G. Pedronette, J. Almeida, R. da S. Torres, A scalable re-ranking method for content-based image retrieval, Inf. Sci. 265 (1) (2014) 91–104.
[73] X. Bai, S. Bai, X. Wang, Beyond diffusion process: neighbor set similarity for fast re-ranking, Inf. Sci. 325 (2015) 342–354.
[74] 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.
[75] 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.
[76] L. Zheng, S. Wang, Q. Tian, Coupled binary embedding for large-scale image retrieval, IEEE Trans. Image Process. (TIP) 23 (8) (2014a) 3368–3380.
[77] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: coupled multi-index for accurate image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2014), 2014b, pp. 1947–1954.
[78] 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.
[79] L. Zheng, S. Wang, Q. Tian, Lp-norm IDF for scalable image retrieval, IEEE Trans. Image Process. 23 (8) (2014) 3604–3617.
[80] B. Wang, J. Jiang, WeiWang, Z.-H. Zhou, Z. Tu, Unsupervised metric fusion by cross diffusion, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2012), 2012, pp. 3013–3020.
[81] L. Xie, R. Hong, B. Zhang, Q. Tian, Image classification and retrieval are one, in: ACM International Conference on Multimedia Retrieval (ICMR), 2015, pp. 3–10.

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 involve content-based image retrieval, unsupervised and semi-supervised learning, image analysis, and digital libraries.

Filipe Marcel Fernandes Gonçalves received a BSc in Life Sciences in (2009) from the State University of São Paulo (Brazil), and in 2015 received a BSc in Computer Science from the same university (UNESP). He received the MSc degree in Mathematics and Computational Intelligence from UNESP in 2016. His research interests involve content-based image retrieval, image analysis, ontology, knowledge engineering and artificial intelligence.

Ivan Rizzo Guilherme received a BSc in Computer Science (1985) from Federal University of São Carlos (Brazil) and the MSc degree in Electrical Engineering (1990) from University of Campinas (Brazil). He got his doctorate in Electrical Engineering at the same university in 1996. He is currently a lecturer at the Department of Statistics, Applied Mathematics and Computing, State University of São Paulo (Brazil). His research interests involve knowledge engineering, semantic Web, intelligent automation and e-learning.