Pattern Recognition Letters 83 (2016) 357–367

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A graph-based ranked-list model for unsupervised distance learning on shape retrieval☆

Daniel Carlos Guimarães Pedronette[a,*], Jurandy Almeida[b], Ricardo da S. Torres[c]

[a] Department of Statistics, Applied Mathematics and Computing, State University of São Paulo (UNESP), Av. 24-A, 1515, Rio Claro, SP 13506-900, Brazil
[b] Institute of Science and Technology, Federal University of São Paulo (UNIFESP), Av. Cesare M. G. Lattes, 1201, São José dos Campos, SP 12247-014, Brazil
[c] Recod Lab, Institute of Computing (IC), University of Campinas (UNICAMP), Av. Albert Einstein, 1251, Campinas, SP 13083-852, Brazil

ARTICLE INFO

Article history:
Available online 26 May 2016

Keywords:
Shape retrieval
Ranking methods
Graph-based approaches

ABSTRACT

Several re-ranking algorithms have been proposed recently. Some effective approaches are based on complex graph-based diffusion processes, which usually are time consuming and therefore inappropriate for real-world large scale shape collections. In this paper, we introduce a novel graph-based approach for iterative distance learning in shape retrieval tasks. The proposed method is based on the combination of graphs defined in terms of multiple ranked lists. The efficiency of the method is guaranteed by the use of only top positions of ranked lists in the definition of graphs that encode reciprocal references. Effectiveness analysis performed in three widely used shape datasets demonstrate that the proposed graph-based ranked-list model yields significant gains (up to +55.52%) when compared with the use of shape descriptors in isolation. Furthermore, the proposed method also yields comparable or superior effectiveness scores when compared with several state-of-the-art approaches.

1. Introduction

Shape matching plays a central role in computer vision and content-based image retrieval systems, as it is one of the most important visual properties in human perception [1–3]. In many recognition applications, for example, the object classes are more easily distinguished using shape features – in opposition to other common properties such as color or texture [4].

Therefore, accurately measuring the similarity between two given shapes represents a fundamental task in many computer vision systems and often depends on an effective shape descriptor, usually defined in terms of a feature extraction function and a similarity measure [5]. During the past decades, several features have been proposed, employing distinct approaches [6].

The contour is a common exploited property, since the object's closed boundary curve contains rich information about the shape complexity [7]. In fact, the contour complexity has been analyzed from different perspectives by different descriptors. The contour saliences, for example, characterizes the influence areas of higher curvature points along a contour [8]. Another strategy relies on the computation of the angle between lines connecting a point with the rest of the points on the boundary [9]. Contour points are also used to compute geometric representations (e.g., triangles) [10].

In addition to contour, another common representation is based on skeletons, which aim at capturing a structural representation of a shape by modeling it in terms of a set of axial curves [11]. Based on the observation that contour-based representations are often effective at representing detailed shape information, and skeleton-based description approaches can cope well with non-rigid deformations, combined approaches also have been proposed [12].

Despite the significant advances achieved by shape descriptors, designing an effective similarity measure has proven to be a challenging task, still considered as a largely unsolved problem [5]. Among the difficulties, we can point out non-linear transformations [13] and the inherent subjectivity associated with the definition of the similarity itself [5].

An innovative perspective to the problem is based on the analysis of the similarity measures in a graph structure [14]. Different from previous advances, which have been mostly driven by designing better shape features, this family of methods learns a new metric through graph transduction by propagation. The Locally Constrained Diffusion Process [1], for example, proposes that shapes do influence the similarity measure of other pairs of shapes, where the influence is propagated as a diffusion process on a graph. Other graph-based initiatives include the Shortest Path Propagation

☆ This paper has been recommended for acceptance by Xiang Bai.

* Corresponding author. Tel.: +55 19 3526 9082.
E-mail addresses: pedronette@gmail.com, daniel@rc.unesp.br, dcarlos@ic.unicamp.br (D.C.G. Pedronette), jurandy.almeida@unifesp.br (J. Almeida), rtorres@ic.unicamp.br (R.d.S. Torres).

http://dx.doi.org/10.1016/j.patrec.2016.05.021
0167-8655/

[15], the Tensor Product Graph [16], and the Self-Smoothing Operator [17].

These graph-based methods present similar objectives: post-processing the distance/similarity measures for improving the effectiveness of retrieval tasks. In general, such methods compute "global" measures capable of taking into account the relationships among images and the structure of the dataset manifold. Some of these methods are represented as specific instances of a generic framework [18] for diffusion processes on an affinity graph. Other variations consider the use of various similarity measures [19,20].

Although indispensable for improving the retrieval effectiveness, the wide use of post-processing methods on large-scale real-world applications also depends on efficiency and scalability aspects [21]. More recently, due to the high computational costs associated with diffusion-based approaches, other efficient post-processing methods have emerged [21–23], mainly based on ranking analysis [24–26].

In these methods, the similarity among ranked lists [24] and the ranking consistency [25] are considered. The reciprocal references also have been attracted a lot of attention [23,26], including fusion tasks [27]. One important advantage of rank-based methods consists in the possibility of processing only a sub-set of ranked lists, reducing the computational costs.

In this paper, we present a novel rank-based algorithm for improving the effectiveness of shape retrieval tasks. The algorithm models each ranked list as a graph, establishing similarity connections among all top-k images. Next, a graph fusion approach is employed for obtaining a single graph representing the whole collection and exploiting the relationships encoded in the dataset manifold. Based on the fused graph, a new distance is learned and a new set of ranked lists is computed.

In summary, the main contribution consists in the proposal of a graph-based model for representing ranked lists. The proposed *Ranked List Graph* considers only the k-neighborhood information, exploiting a recent rank correlation measure [28]. In this way, computational costs are restricted only to the top-k positions, providing, at the same time, an effective and efficient representation for search tasks. Additionally, the method requires a very small number of iterations to obtain the best effectiveness results.

Extensive experiments were conducted on a three public datasets and considering different descriptors. Experimental results demonstrate that the proposed method can obtain significant effectiveness gains (up to +55.52% in terms of relative gains). Our approach was also evaluated in comparison with other state-of-the-art approaches, yielding effectiveness results superior and comparable to various post-processing algorithms recently proposed in the literature. For example, we have achieved 100% accuracy (P@20) on the well-known MPEG-7 shape dataset.

The paper is organized as follows: in Section 2, a formal definition of the addressed problem is discussed. Section 3 presents the proposed graph-based distance learning approach. Section 4 discusses the experimental evaluation and, finally, Section 5 draws our conclusions and presents future work.

2. Problem formulation

A formal definition of the image retrieval model considered is presented in this section. Let $\mathcal{C}=\{img_1, img_2, \ldots, img_n\}$ be an image collection. Let $n = |\mathcal{C}|$ be the size of the collection $\mathcal{C}$. Let $\mathcal{D}$ be an image descriptor, which can be defined according to [29] 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 $\hat{I}$; 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.

The distance between two images img_i and img_j is defined by the value of $\rho(\epsilon(img_i), \epsilon(img_j))$. However, the notation $\rho(i, j)$ is used along the paper for readability purposes.

Based on the distance function ρ, a ranked list τ_q can be computed in response to a query image img_q. The ranked list τ_q=(img_1, img_2, …, img_n) can be defined as a permutation of the collection $\mathcal{C}$. A permutation τ_q is a bijection from the set $\mathcal{C}$ onto the set $[N] = \{1, 2, \ldots, n\}$. The value of $\tau_q(i)$ can be interpreted as the position (or rank) of image img_i in the ranked list τ_q. In other words, if img_i is ranked before img_j in the ranked list of img_q (that is, $\tau_q(i) < \tau_q(j)$), then $\rho(q, i) \leq \rho(q, j)$.

Every image $img_q \in \mathcal{C}$ can be taken as a query image in order to compute a ranked list for each image of the collection. In this way, a set of ranked lists $\mathcal{R} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained. The unsupervised distance learning aims at exploiting the information encoded in the set of ranked lists $\mathcal{R}$ for computing a more effective distance function ρ_c. Subsequently, a new set of ranked lists $\mathcal{R}_c$ can be computed based on distance ρ_c. More formally, we can define the unsupervised distance learning algorithm as a function f_r:

$$\mathcal{R}_c = f_r(\mathcal{R}). \tag{1}$$

Additionally to the objective of improving the effectiveness of the retrieval results, efficiency aspects are also considered. In this way, the algorithm processes only sub-sets of the ranked lists, with fixed sizes denoted by constants k and L. The most relevant information are expected to be at the top-k positions of ranked lists, but useful information can be obtained until position L, such $k \leq L \ll n$.

3. Ranked-list graph model

In this section, we present the Ranked-List Graph Distance algorithm for distance learning and distance fusion tasks.

3.1. Distance learning

The main contribution of the proposed algorithm consists in modeling each ranked list as a graph, considering its top-k positions. Different from pairwise distances, which consider the query image and each image in isolation, the proposed ranked list graph establishes relationships among the query and all its neighbors.

Once each collection image and its respective ranked list is represented by a graph, a fusion approach is employed to combine them into a single graph representing the whole collection. Finally, the learned distance is computed based on this graph.

The graph-based approach can be roughly divided into four steps:

1. **Rank normalization:** this step is in charge of recomputing the distances among images by considering their mutual reference defined by their respective ranked lists;
2. **Ranked-List Graph Computation:** this step constructs a graph for representing top-k positions of ranked lists;
3. **Graph fusion:** this step combines all ranked-list graphs into a single collection graph;
4. **Ranked-List Graph Distance Computation:** this step computes the *Ranked-List Graph Distance* among images based on the edges of the collection graph.

Fig. 1 illustrates in a simplified form the graph-based approach for modeling the ranked lists. Given two images img_i, img_j, their respective ranked lists τ_i, τ_j are represented as two graphs.

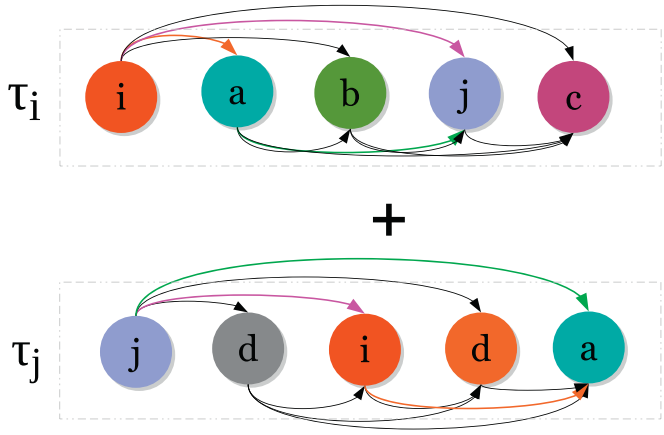

Fig. 1. Overview of the proposed Ranked-List Graph model.

Observe the ranked list τ_i, for example: beyond the edges among the query image (img_i) and the top retrieved results (img_a, img_b, img_j, img_c), the graph also connects the top retrieved results themselves. In the following, the fusion of graphs defined by distinct ranked lists reinforces recurring edges, as occurs for edges among img_i, img_j, img_a (in colors).

Next, a formal definition of each step of the algorithm is presented.

3.1.1. Rank normalization

While most of similarity/dissimilarity pairwise measures are symmetric, the same does not occur for rank analysis. In this way, an image img_i well ranked for a query img_j, does not imply that img_j is well ranked for query img_i. The benefits from improving the symmetry of the k-neighborhood relationship are well known [30] for image retrieval.

In this work, a simple approach, which considers the mutual reference among ranked lists, is employed. Only the images at the top-L positions of the ranked lists are considered, aiming at keeping the low computational costs. A rank normalized distance ρ_n is computed as:

$$\rho_n(i,j) = \tau_i(j) + \tau_j(i), \tag{2}$$

where $\tau_i(j) \leq L$. In the following, all the ranked lists are updated according to the rank normalized distance ρ_c, defining a new set of ranked lists $\mathcal{R}_n$, which is used in the next steps of the algorithm.

3.1.2. Ranked-List Graph Computation

The Ranked-List Graph is defined in terms of the k-neighborhood of collection images.

Let $\mathcal{N}(i,k)$ be the neighborhood set, which is formally defined as follows:

$$\mathcal{N}_k(i) = \{\mathcal{R} \subseteq \mathcal{C}, |\mathcal{R}| = k \wedge \forall x \in \mathcal{R}, y \in \mathcal{C} - \mathcal{R} : \rho_n(i,x) \leqslant \rho_n(i,y)\} \tag{3}$$

The Ranked-List Graph computed for an image img_i is defined as a weighted undirected graph $G_i = (V_i, E_i, w_i)$, where the set of vertices V_i is defined by the set $\mathcal{N}_k(i)$. Each image is represented by a node and $V_i = \mathcal{N}_k(i)$. The edge set E_i is defined considering the correlation among images at the top n_s positions of each ranked list, as follows:

$$E_i = \{(img_j, img_l) \mid img_j, img_l \in \mathcal{N}_k(i)\}. \tag{4}$$

The edge weight w_i is defined by a recently proposed rank measure [28], based on a probabilistic user model employed for performing rank correlation analysis. The Rank-Biased Overlap [28] (RBO) compares the overlap of two ranked lists at incrementally increasing depths. The measure takes a parameter p that specifies the probability of considering the overlap at the next level. The RBO measure is formally defined as follows:

$$RBO(j,l) = (1-p)\sum_{d=1}^{k} p^{d-1} \times \frac{|\mathcal{N}_k(j) \cap \mathcal{N}_k(l)|}{d}, \tag{5}$$

where p is a constant, which determines the strength of the weighting to top ranks. The edge weight w_i is defined by the RBO measure, such that $w_i(j,a) = RBO(j,a)$.

3.1.3. Graph fusion

Recently, graph fusion approaches have been proposed for combining different retrieval methods [27]. In this work, we employ a fusion method for combining the graphs of different ranked lists into a novel representation named Collection Graph. The Collection Graph represents the relationships among all images in a collection, through the fusion of Ranked-List Graphs.

Despite the use of a graph-based model, the proposed method differs from [27] in many aspects: the graph defined by Zhang et al. [27] has an edge only if two images are reciprocal neighbors. In the proposed Ranked-List Graph, the edges connect the query image and all their neighbors, defining a fully connected graph for each ranked list. Such methods also differ regarding the weight of the edges: the Jaccard and a decay coefficient related to the number of hops to the query is used in [27], while our approach uses the RBO [28] measure. Additionally, while [27] performs a ranking step using a transition matrix based on PageRank or a greedy algorithm, the proposed approach does not require any analogous step.

The proposed Collection Graph G_c is defined as a weighted undirected graph $G_c = (V_c, E_c, w_c)$, where the set of vertices V_c is defined by the image collection $\mathcal{C}$, such that $V_c = \mathcal{C}$. The edge set E_c is defined as the union of edges defined for all ranked list graphs. Formally, we have $E_c = \bigcup_i E_i$ for each image $img_i \in \mathcal{C}$.

The edge weights w_c are also defined in terms of the sum of all weights defined in the ranked-list graphs, as:

$$w_c(j,l) = \sum_{img_i \in \mathcal{C}} w_i(j,l). \tag{6}$$

Notice that the edge weight, which defines the strength of connection (and the similarity) between images img_j and img_l, is computed based on information encoded in the weights of all ranked lists (w_i, with each $img_i \in \mathcal{C}$).

3.1.4. Ranked-List Graph Distance

Based on the weights of the Collection Graph, a new distance ρ_c can be computed. Given two images img_i, img_j, the distance between them $\rho_c(i,j)$ is defined as follows:

$$\rho_c(i,j) = \frac{1}{1 + w_c(i,j)}, \tag{7}$$

where $w_c(i,j) > 0$. For images without edges, the ranked lists remain the same. More formally, for img_i, img_l such $(img_i, img_l) \cap E_c = \{\emptyset\}$, we have $\rho_c(i,l) = \tau_i(l)$.

A new set of ranked lists $\mathcal{R}_c$ is computed based on the new distance ρ_c. Once the input for the ranked list graphs consists in the set of ranked lists, the process can iteratively be repeated along iterations. Let $^{(t)}$ denotes the current iteration, an iterative distance function $\rho_c^{(t)}(i,j)$ can be defined, and consequently, a set of ranked lists $\mathcal{R}_c^{(t)}$.

The final set of ranked lists $\mathcal{R}_c^{(T)}$ is obtained after T iterations of the algorithm. As discussed in the experimental section, the number of required iterations is very small. In fact, the most significant gains are obtained at the first iteration.

Algorithm 1 Rank normalization algorithm.

Require: Set of ranked lists $\mathcal{R}$, parameter L
Ensure: Rank Normalized Set of Ranked Lists $\mathcal{R}_n$
1: **for all** $img_i \in \mathcal{C}$ **do**
2: **for all** $img_j \in \mathcal{N}_L(i)$ **do**
3: $\rho_n(i,j) \leftarrow \tau_i(j) + \tau_j(i)$
4: **end for**
5: **end for**
6: $\mathcal{R}_n = sort(\mathcal{R}, \rho_n)$

Algorithm 2 Rank List Graph Distance algorithm.

Require: Rank Normalized Set of Ranked Lists $\mathcal{R}_n$, parameters k, T
Ensure: Updated set of Ranked Lists $\mathcal{R}_c$
1: $t \leftarrow 0$
2: $\mathcal{R}_c^{(0)} \leftarrow \mathcal{R}_n$
3: **while** $t < T$ **do**
4: **for all** $img_i \in \mathcal{C}$ **do**
5: **for all** $img_j \in \mathcal{N}_k(i)$ **do**
6: **for all** $img_l \in \mathcal{N}_k(i)$ **do**
7: $w_c(j,l) \leftarrow RBO(j,l)$
8: **end for**
9: **end for**
10: **end for**
11: **for all** $img_i \in \mathcal{C}$ **do**
12: **for all** $img_j \in \mathcal{C}$ **do**
13: $\rho_c^{(t)} \leftarrow 1/(1 + w_c(i,j))$
14: **end for**
15: **end for**
16: $\mathcal{R}_c^{(t+1)} = sort(\mathcal{R}_c^{(t)}, \rho_c^{(t)})$
17: $t \leftarrow t + 1$
18: **end while**

3.1.5. Rank-List Graph algorithm

This section discusses an algorithmic solution for the proposed method. Algorithm 1 presents an approach for computing the rank normalization step. The distance ρ_n is updated (Line 3) according to the top-L positions defined by the set $\mathcal{N}_L$ (Line 2).

Algorithm 2 outlines the method for updating the set of ranked lists according to the Ranked-List Graph Distance. For each image img_i (Line 4) and its neighbors img_j and img_l (Lines 5–6), the weights of edges from Ranked-List Graph are added to the Collection Graph weights w_c (Line 7). The new distance is computed in Lines 11-15, leading to an updated set of ranked lists (Line 16).

For a given image collection with n images, the most important steps of the proposed method are restricted to the top-k or top-L positions of the n ranked lists. Algorithm 2 (Lines 11–14) redefines the distance among all images, but it can be easily adapted to recompute only the distances until the top-L positions of each ranked list, similarly to Algorithm 1. The same can be considered for the sorting step. In this way, with all operations restricted to top-L positions, the overall algorithm presents a complexity of only $O(n)$.

3.2. Descriptor combination

Different image descriptors may focus on diverse and complementary aspects of the shape, like contour, curvature, and skeleton. Therefore, it is intuitive that the combination of the distances computed according to different features can improve the retrieval accuracy [31].

In this work, we exploit the Ranked-List Graph Distance in conjunction with a multiplicative approach inspired on recent positive results [23,32] for combining image descriptors. First, the Ranked-List Graph Distance is computed in isolation for each feature, considering one iteration. Subsequently, the results are multiplied and combined into a single distance. Besides being unsupervised, the proposed method does not perform any normalization steps as required by other fusion approaches [31].

Let $\rho_{c_d}^{(1)}$ be the Ranked-List Graph Distance at first iteration of a given descriptor d and let m be the number of descriptors considered, the fused distance can be defined as:

$$\rho_f^{(1)}(i,j) = \prod_{d=1}^{m}(1 + \rho_{c_d}^{(1)}(i,j)). \quad (8)$$

Once a combined distance is computed, a set of ranked lists $\mathcal{R}_f^{(1)}$ can be obtained and other iterations of the algorithm can be processed. After T iterations of the algorithm, the final set of ranked lists $\mathcal{R}_f^{(T)}$ is obtained. As discussed in experimental section, for distance fusion tasks only one iteration ($T = 1$) is needed for the combined distance.

4. Experimental evaluation

In this section, we present the results of experimental evaluation conducted for assessing the effectiveness of the proposed method. A rigorous experimental protocol was employed, involving three different datasets and various descriptors. Our objective is to evaluate the proposed method in diverse scenarios. We also performed statistical tests to confirm if the use of the proposed post-processing approach yields significant results, when compared to the use of the shape descriptors in isolation.

4.1. Datasets and descriptors

The datasets and descriptors considered in the experimental evaluation are briefly described in this section.

4.1.1. MPEG-7

The MPEG-7 [35] dataset is a well-known shape dataset, composed of 1,400 shapes which are grouped into 70 classes, with 20 objects per class. The dataset is widely used for shape retrieval and post-processing methods evaluation.

Six different shape descriptors are considered: Segment Saliences (SS) [3], Beam Angle Statistics (BAS) [9], Inner Distance Shape Context (IDSC) [13], Contour Features Descriptor (CFD) [33], Aspect Shape Context (ASC) [36], and Articulation-Invariant Representation (AIR) [34].

Two effectiveness measures were considered the for the MPEG-7 [35] dataset: the MAP and the Bull's Eye Score, commonly used for this dataset. The score counts all matching shapes within the top-40 ranked images. The retrieved score is normalized, since each class consists of 20 shapes which defines highest possible number of hits, being equivalent to Recall@40. For data fusion, the accuracy score which is a more strict measure equivalent to the P@20 is also considered.

4.1.2. Animal

The Animal[1] [12] dataset is composed of 2,000 animal shapes from 20 different classes. This dataset is equally divided into two parts labeled A and B, and each part comprises 10 classes of animals. Each class contains 100 shape images from different animal views.

[1] https://sites.google.com/site/xiangbai/animaldataset (As of November 2015).

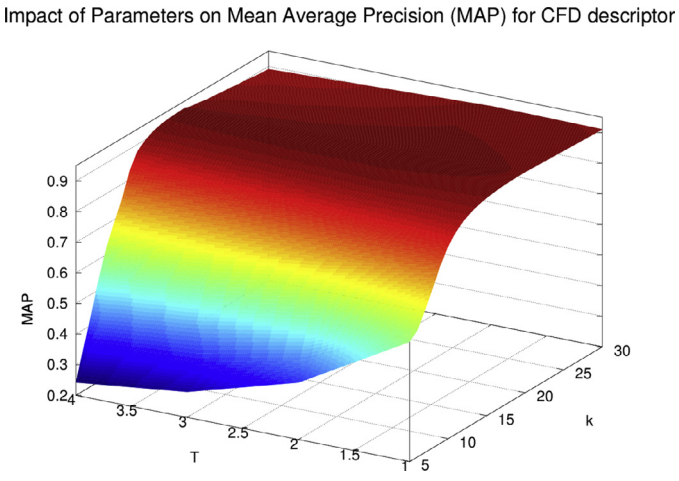

Fig. 2. Impact of parameters k and T for the CFD [33] descritor.

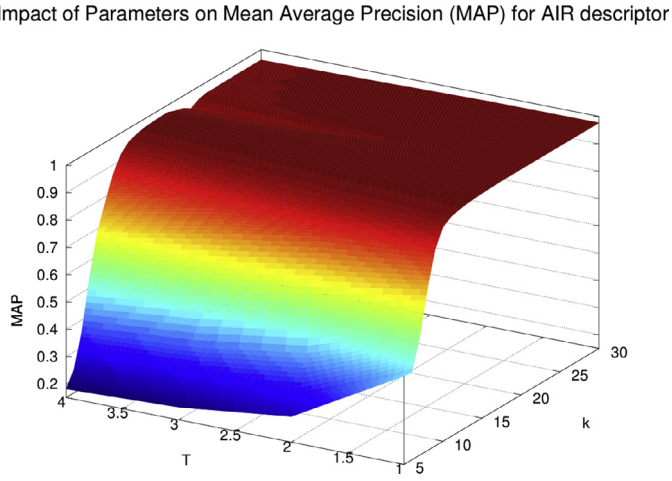

Fig. 3. Impact of parameters k and T for the AIR [34] descritor.

The descriptors used for encoding shape properties are: Fourier Descriptor (FD) [37], Curvature Scale Space (CSS) [38], Tensor Scale Descriptor (TSD) [39], Segment Saliences (SS) [3], Beam Angle Statistics (BAS) [9], and Triangle Area Representation (TAR) [40]. The effectiveness of each descriptor was assessed using three metrics: MAP, P@10, and P@20.

4.1.3. ETH-80

The ETH-80[2] [41] dataset is composed of 3,280 images, and each image comprises one single object at its center and a known background. This dataset consists of 80 objects from 8 different classes. Each class contains 10 objects with 41 view per object.

For each object, there is provided a color image and a binary image of its contour, which we used in our shape experiments. All the images are cropped and rescaled to a size of 128 × 128 pixels. The shape descriptors and effectiveness measures considered for the ETH-80 dataset were the same used for the Animal dataset.

4.2. Impact of parameters

This section aims at assessing the robustness of the method to different parameter settings, evaluating the impact different parameter values on the effectiveness results. We conducted various experiments considering the MPEG-7 collection [35].

The first experiment evaluates the impact of the parameters k (size of the neighborhood set) and T (number of iterations). Figs. 2 and 3 illustrate the effectiveness scores given by the Mean Average Precision (MAP) according to variations of k and T, for descriptors CFD [33] and AIR [34], respectively. A large and stable red region can be observed for both surfaces, demonstrating the robustness of the method in achieving high effectiveness gains for different parameters settings. The best effectiveness results are obtained by values near to $k = 20$ and $T = 2$. In most of remaining experiments, we used $k = 20$ as the neighborhood size, except for the Animal dataset, in which we used $k = 40$ due to the larger number of images within each class.

The impact of the size of subset of ranked lists (L) is also evaluated, considering three different shape descriptors: CFD [33], ASC [36], and AIR [34]. Fig. 4 shows the impact of this parameter on the MAP scores. A fast growth of effectiveness scores can be observed for small values of L. The effectiveness gains stabilize for the three descriptors for values between 100 and 200. For most of experiments, we used L=100.

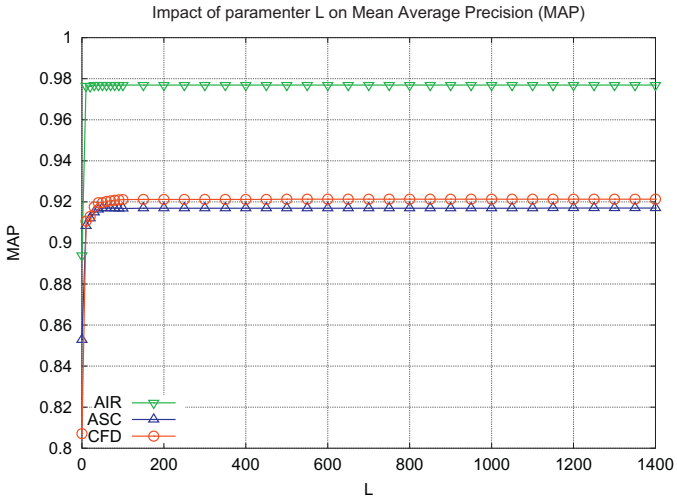

Fig. 4. Impact of parameter L on effectiveness.

The last experiment evaluated the parameter p used for the RBO measure. Fig. 5 shows the variation of MAP scores according to different values of p. The parameter p indicates the weight given to top positions of ranked lists. The descriptors CFD [33], ASC [36], and AIR [34] are considered. The results varied, according to the effectiveness of the descriptors. For the AIR descriptor, which presents higher effectiveness scores, greater values of p presented better results. We used $p = 0.95$ in the remaining experiments.

4.3. Evaluation of the proposed method

Various experiments were conducted for evaluating the effectiveness of the proposed method, considering the three datasets and image descriptors discussed in Section 4.1. A statistical analysis is also presented, using paired t-tests. The objective consists in assessing if the difference between the retrieval results before and after the use of the algorithm is statistically significant.

4.3.1. MPEG-7 dataset

The first experiment considered the MPEG-7 dataset and the Bull's Eye Score (Recall@40). Table 1 presents the effectiveness results and the relative gains for six image descriptors. The full ranked lists (L=1400) were considered and different values of iterations ($T = 1$ and $T = 2$). Very significant positive gains can be observed, ranging from +6.81% to +35.24%.

[2] http://www.mis.informatik.tu-darmstadt.de/Research/Projects/categorization/eth80-db.html (As of August 2007).

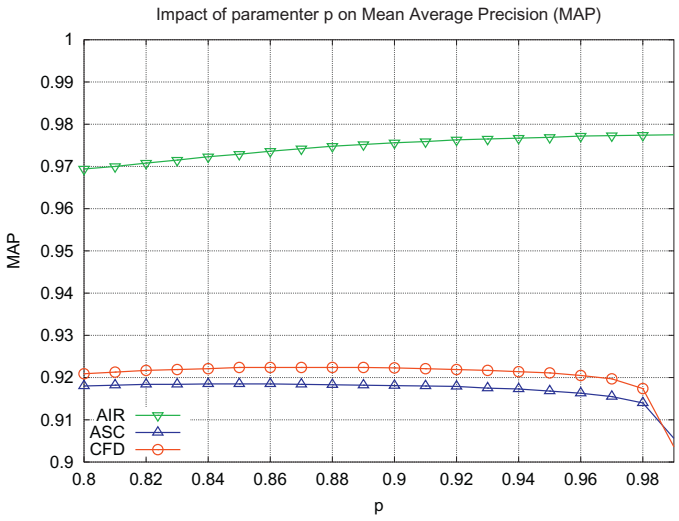

Fig. 5. Impact of parameter p (RBO measure) on effectiveness.

An analogous experiment was conducted aiming at evaluating the impact of considering only top positions of ranked lists ($L = 100$). Table 2 presents the results. As we can observe, the effectiveness results are very similar to full ranked lists, demonstrating the capability of the method in achieving effectiveness gains without neglecting efficiency aspects. All remaining experiments considered $L = 100$.

The proposed algorithm is also evaluated on the MPEG-7 dataset considering the MAP scores. Table 3 presents the obtained results. We can observe that the relative gains obtained for MAP are even greater than for Recall@40. For the SS [3] descriptor, for example, the Ranked-List Graph distance improved the results from 37.67% to 52.51%, achieving a relative gain of +39.39%. Notice also that all the results for the MPEG-7 dataset are statistically significant at a confidence of 99%.

4.3.2. Animal dataset

The experimental results for the Animal dataset, Parts A and B, are presented in Tables 4 and 5, respectively. The MAP measure is considered for this experiment. For the Part A of the dataset, only positive gains are obtained, ranging from +1.18% to 45.57%. For the Part B, the algorithm achieved even more impressive gains, reaching +55.52% for the BAS [9] descriptor.

The only exception is the FD descriptor, where no gains are observed. It is worth mentioning that the considered descriptors for the Animal dataset (and also for the ETH-80 collection) achieved a lower effectiveness scores, if compared with the MPEG-7 dataset. This scenario is more challenging for unsupervised algorithms, which depends on the existence of relevant results in the top-ranked positions.

4.3.3. ETH-80 dataset

Table 6 presents the MAP scores for the ETH-80 dataset. Despite the low initial effectiveness scores (which are even smaller than the Animal dataset), the algorithm achieved significant gains, except for the FD descriptor. For example, the algorithm improved

Table 1
Ranked List Graph Distance on the MPEG-7 dataset, considering the Bull's Eye Score (Recall@40) and $L = 1400$.

Shape descriptor	Original Bull's Eye Score (%)	Ranked List Graph Dist. $T = 1$ (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. $T = 2$ (%)	Gain (%)	Statistic signific. 99%
SS [3]	43.99	57.61	+30.94	•	59.55	+35.35	•
BAS [9]	75.20	86.14	+14.53	•	87.35	+16.15	•
IDSC [13]	85.40	92.15	+7.75	•	92.90	+8.62	•
CFD [33]	84.43	94.97	+12.47	•	95.66	+13.29	•
ASC [36]	88.39	94.42	+6.82	•	95.19	+7.71	•
AIR [34]	93.67	99.99	+6.88	•	99.59	+6.45	•

Table 2
Ranked List Graph Distance on the MPEG-7 dataset, considering the Bull's Eye Score (Recall@40) and $L = 100$.

Shape descriptor	Original Bull's Eye Score (%)	Ranked List Graph Dist. **T = 1** (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. **T = 2** (%)	Gain (%)	Statistic signific. 99%
SS [3]	43.99	57.51	+30.71	•	59.51	+35.24	•
BAS [9]	75.20	85.83	+14.12	•	87.19	+15.93	•
IDSC [13]	85.40	92.14	+7.74	•	92.89	+8.61	•
CFD [33]	84.43	94.96	+12.46	•	95.66	+13.29	•
ASC [36]	88.39	94.41	+6.81	•	95.19	+7.71	•
AIR [34]	93.67	99.99	+6.88	•	99.59	+6.45	•

Table 3
Ranked List Graph Distance on the MPEG-7 dataset, considering the MAP scores and $L = 100$.

Shape descriptor	Original MAP score (%)	Ranked List Graph Dist. **T = 1** (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. **T = 2** (%)	Gain (%)	Statistic signific. 99%
SS [3]	37.67	49.97	+32.65	•	52.51	+39.39	•
BAS [9]	71.52	81.15	+13.46	•	82.58	+15.46	•
IDSC [13]	81.70	89.20	+9.18	•	90.17	+10.37	•
CFD [33]	80.71	92.11	+14.12	•	93.25	+15.54	•
ASC [36]	85.28	91.68	+7.50	•	92.69	+8.69	•
AIR [34]	89.39	97.69	+9.29	•	97.36	+8.92	•

Table 4
Ranked List Graph Distance on the Animal-A dataset, considering the MAP scores and $L = 100$.

Shape descriptor	Original MAP score (%)	Ranked List Graph Dist. **T = 1** (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. **T = 2** (%)	Gain (%)	Statistic signific. 99%
FD	13.62	13.84	+1.59	•	13.78	+1.18	•
CSS	14.57	15.09	+3.57	•	14.78	+1.44	
TSD	20.66	21.90	+5.96	•	21.80	+5.50	•
SS	23.75	30.02	+26.40	•	32.53	+36.93	•
BAS	27.61	38.22	+38.40	•	40.20	+45.57	•
TAR	34.73	42.56	+22.56	•	44.38	+27.80	•

Table 5
Ranked List Graph Distance on the Animal-B dataset, considering the MAP scores and $L = 100$.

Shape descriptor	Original MAP score(%)	Ranked List Graph Dist. **T = 1** (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. **T = 2** (%)	Gain (%)	Statistic signific. 99%
FD	15.23	15.14	−0.63		15.00	−1.54	•
CSS	14.91	15.67	+5.08	•	15.32	+2.75	
TSD	17.18	17.60	+2.42	•	17.60	+2.42	•
SS	22.18	29.83	+34.52	•	32.92	+48.47	•
BAS	28.87	43.38	+50.29	•	44.89	+55.52	•
TAR	43.05	50.55	+17.44	•	51.89	+20.54	•

Table 6
Ranked List Graph Distance on the ETH-80 dataset, considering the MAP scores and $L = 100$.

Shape descriptor	Original MAP score (%)	Ranked List Graph Dist. **T = 1** (%)	Gain (%)	Statistic signific. 99%	Ranked List Graph Dist. **T = 2** (%)	Gain (%)	Statistic signific. 99%
FD	7.99	7.81	−2.29	•	7.71	−3.39	•
CSS	5.90	7.03	+19.26	•	6.72	+13.90	•
SS	10.98	14.15	+28.78	•	15.04	+36.89	•
BAS	12.65	19.21	+51.89	•	19.48	+54.05	•
TSD	14.82	16.51	+11.45	•	16.63	+12.23	•
TAR	19.84	21.01	+5.83	•	20.88	+5.22	•

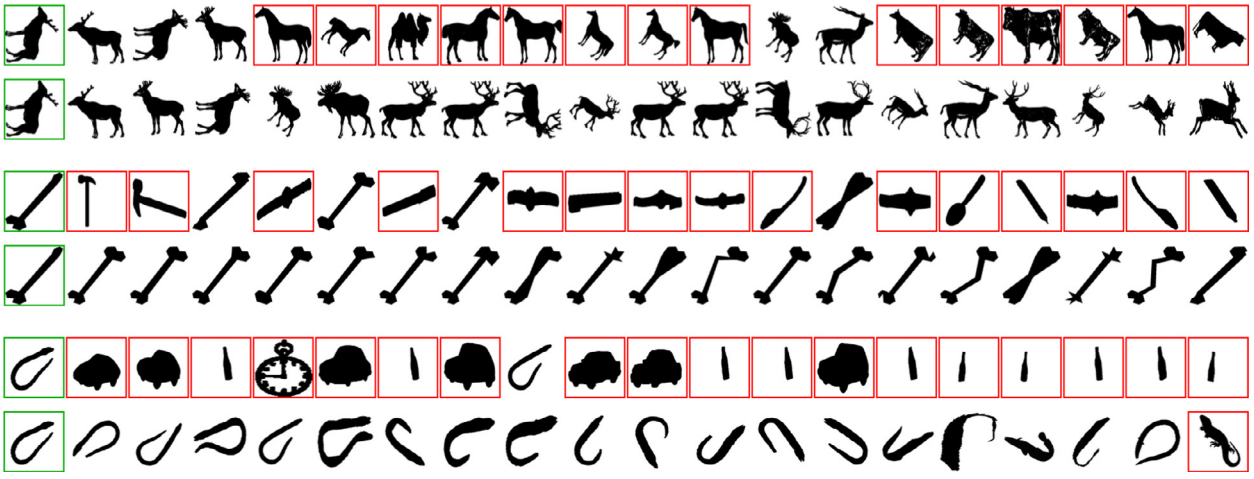

Fig. 6. Visual examples of retrieval results before and after the use of the Ranked-List Graph algorithm, considering the CFD descriptor: query image with green border and wrong images with red borders. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the MAP score of the BAS descriptor from 12.65% to 19.48%, reaching a relative gain of +54.05%.

4.4. Visual re-ranked results

This section aims at illustrating the visual impact of the proposed algorithm on retrieval results. Fig. 6 illustrates the returned ranked lists for the CFD [33] descriptor on MPEG-7. Three query images are illustrated before and after the execution of the algorithm. The first image of each row represents the query image (highlighted with a green border). Wrong (non-relevant) images in the ranked lists are illustrated with red borders.

The effectiveness of retrieval is greatly improved for all the queries, from 25% and 20% to 100% for the first and second queries, and from 5% to 95% for the third query. The third query is a remarkable example of the capacity of the algorithm in exploiting the information encoded in the whole dataset. Despite the absence of correct enough information in the ranked list, other ranked lists

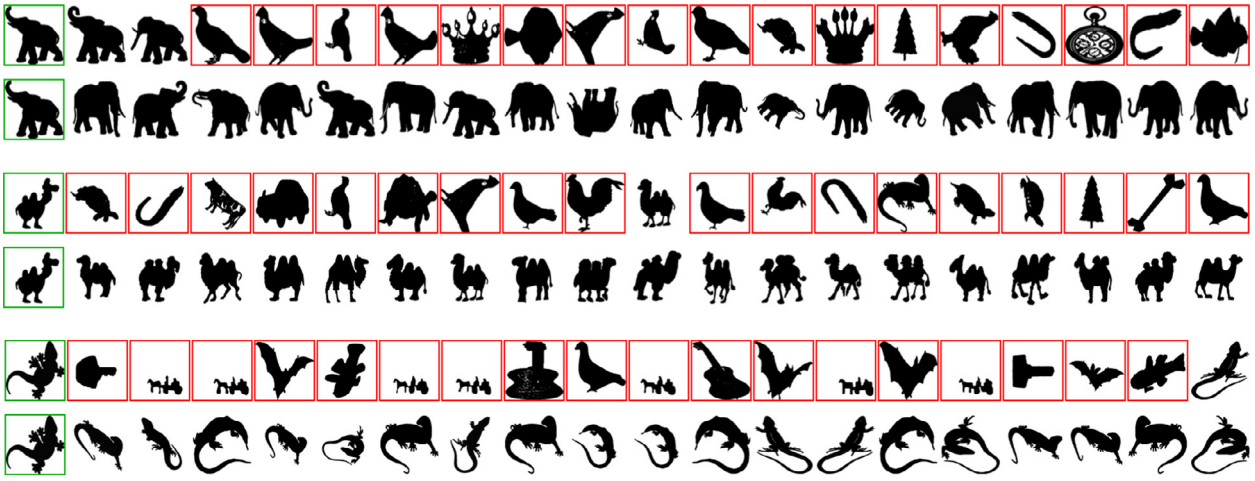

Fig. 7. Visual examples of retrieval results before and after the use of the Ranked-List Graph algorithm, considering the ASC descriptor: query image with green border and wrong images with red borders. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
Distance fusion by **Ranked-List Graph** on the MPEG-7 dataset, considering different retrieval scores: Bull's Eye Score (Recall@40), MAP, and Accuracy (P@20).

Descriptor	Bull's Eye Score (%)	MAP (%)	Accuracy (P@20) (%)
SS [3]	43.99	37.67	35.75
BAS [9]	75.20	71.52	67.22
IDSC [13]	85.40	81.70	77.21
CFD [33]	84.43	80.71	75.59
ASC [36]	88.39	85.28	80.66
AIR [34]	93.67	89.39	88.17
SS+BAS	86.41	81.17	77.14
SS+IDSC	96.15	94.01	91.98
BAS+IDSC	96.76	95.27	93.26
CFD+ASC	99.62	99.06	98.28
CFD+AIR	100	100	100
ASC+AIR	99.92	99.75	99.47

are considered allowing the improvement of retrieval results. Fig. 7 presents ranked lists considering the ASC [36] descriptor. Again, similar positive results are observed.

4.5. Evaluation of the descriptor combination

The experimental results of the Ranked-List Graph on distance fusion tasks are discussed in this section. For the MPEG-7 dataset, we considered two groups of descriptors, according to the effectiveness achieved in distance learning tasks: SS, BAS, and IDSC, as the first group; and CFD, ASC, and AIR as the second group. All the combinations among descriptors in each group are evaluated. In addition to the MAP and Bull's Eye scores, the Accuracy measure is also reported considering the precision of retrieval at top-20 positions.

Table 7 presents the results obtained for the MPEG-7 dataset. For all combinations and effectiveness measures, the combined result is better than the use of the best descriptor in isolation. The BAS+IDSC combination, for instance, achieved an accuracy score of 93.26%, while the scores of descriptors in isolation are only 67.22% and 77.21%, respectively. The CFD+AIR combination, in turn, achieved 100% for the three measures, indicating perfect retrieval results.

The Ranked-List Graph also achieved similar positive results for other datasets. Tables 8 and 9 present the results for Animal dataset, parts A and B, respectively. The MAP and precision measures are evaluated considering the three descriptors with the best

Table 8
Distance fusion by **Ranked-List Graph** on the Animals-A dataset.

Descriptor	P@10 (%)	P@20 (%)	MAP (%)
SS	47.77	39.45	23.75
BAS	49.57	48.05	27.61
TAR	69.46	60.43	34.72
SS+BAS	66.30	60.79	42.80
SS+TAR	69.73	64.54	45.86
BAS+TAR	71.95	65.83	44.81

Table 9
Distance fusion by **Ranked-List Graph** on the Animals-B dataset.

Descriptor	P@10(%)	P@20(%)	MAP(%)
SS	43.16	35.49	22.18
BAS	45.20	47.17	28.87
TAR	70.39	63.28	43.05
SS+BAS	70.49	65.79	46.82
SS+TAR	73.52	69.19	52.53
BAS+TAR	72.99	68.45	51.64

Table 10
Distance fusion by **Ranked-List Graph** on the ETH-80 dataset.

Descriptor	P@10 (%)	P@20 (%)	MAP (%)
BAS	37.62	29.02	12.64
TSD	35.28	26.20	14.82
TAR	41.26	31.60	19.85
BAS+TSD	41.15	32.13	21.68
BAS+TAR	39.97	31.12	21.00
TSD+TAR	42.35	33.28	22.64

effectiveness scores on distance learning tasks. We may highlight, for instance, the improvements obtained for the SS+BAS combination on part B, from initial MAP scores of 22.18% and 28.87% to a combined score of 46.82%.

The results of ETH-80 dataset are presented in Table 10. Again, the combined results are better than the best descriptor in isolation for all combinations. Despite the positive results, the gains are smaller than those obtained of other datasets, mainly due to the lower initial effectiveness scores of this dataset.

Table 11
Comparison with post-processing methods on the MPEG-7 dataset, considering the *Bull's Eye Score* (Recall@40).

Algorithm	Descriptor(s)	Bull's eye Score (%)
Shape descriptors		
DDGM [42]	–	80.03
CFD [33]	–	84.43
IDSC [13]	–	85.40
SC [43]	–	86.80
ASC [36]	–	88.39
AIR [34]	–	93.67
Unsupervised post-processing methods: distance learning		
Graph transduction [14]	IDSC	91.61
LCDP [1]	IDSC	93.32
Shortest Path Propagation [15]	IDSC	93.35
Mutual kNN graph [44]	IDSC	93.40
Pairwise recommendation [32]	ASC	94.66
RL-Sim [24]	ASC	94.69
Ranked List Graph Dist.	**ASC**	**95.19**
Ranked List Graph Dist.	**CFD**	**95.66**
LCDP [1]	ASC	95.96
Tensor Product Graph [16]	ASC	96.47
RL-Sim [24]	AIR	99.94
Reciprocal kNN manifold [23]	AIR	99.94
Tensor Product Graph [16]	AIR	99.99
Ranked List Graph Dist.	**AIR**	**99.99**
Generic diffusion process [18]	AIR	100
Neighbor set similarity [22]	AIR	100
Unsupervised post-processing methods: distance fusion		
Reciprocal rank fusion [45]	CFD+IDSC	94.98
Graph fusion [27]	CFD+ASC	96.16
Reciprocal rank fusion [45]	CFD+ASC	96.25
Co-transduction [20]	SC+DDGM	97.45
Self-smoothing operator [17]	SC+IDSC	97.64
Co-transduction [20]0	SC+IDSC	97.72
Self-smoothing operator [17]	SC+IDSC+DDGM	99.20
Pairwise recommendation [32]	CFD+IDSC	99.52
Ranked List Graph Dist.	**CFD+ASC**	**99.62**
RL-Sim [24]	CFD+ASC	99.65
Ranked List Graph Dist.	**CFD+AIR**	**100**

Table 12
Comparison on the MPEG-7 dataset, considering *Accuracy score* (P@20).

Unsupervised post-processing methods: distance fusion		
Algorithm	Descriptor(s)	Accuracy (%)
Co-transduction [20]	IDSC+DDGM	95.12
Co-transduction [20]	SC+IDSC+DDGM	95.24
Cross diffusion process [19]	IDSC+DDGM	99.69
Cross diffusion process [19]	SC+IDSC	99.86
Cross diffusion process [19]	SC+IDSC+DDGM	100
Reciprocal kNN distance [23]	CFD+AIR	100
Ranked List Graph Dist.	**CFD+AIR**	**100**

4.6. Comparison with other approaches

The Ranked List Graph Distance was also evaluated in comparison with various state-of-the-art methods. The MPEG-7 dataset was considered due to its frequent use for evaluation and comparison among post-processing methods. An experimental protocol commonly reported in the literature was followed, using the Bull's Eye Score as effectiveness measure and all images as queries. Table 11 presents the best results of the proposed algorithm (in bold) in comparison with several other methods on distance learning and fusion tasks. Despite the small sub-set of ranked lists required, the proposed approach achieved high effectiveness scores, comparable and better than various recently proposed methods.

Due to the saturation of the Bull's Eye Score, we also considered the Accuracy score for evaluation on distance fusion tasks. Table 12 presents the results of proposed method in comparison with state-of-the-art approaches. The Ranked List Graph distance achieved an Accuracy score of 100% for fusion of CFD+AIR. The Accuracy score is a stricter measure than the Bull's Eye Score, and 100% indicates perfect retrieval results, achieved only by few methods.

Table 13
Ranked List Graph Distance on the Holidays dataset.

Descriptor	Original MAP (%)	Ranked List Graph Dist. (%)	Relative gain (%)
JCD [49]	52.83	55.04	+4.18
SCD [48]	54.26	56.60	+4.31
ACC [47]	64.29	70.37	+9.46
CNN-Caffe [50]	64.09	70.78	+10.44
CNN-OverFeat [51]	82.59	**85.33**	+3.32
ACC + OverFeat	–	82.71	+28.65
ACC + Caffe	–	77.84	+21.45
ACC + Caffe + OverFeat	–	84.33	+31.58

Table 14
Comparison with state-of-the-art on the Holidays dataset.

MAP scores for recent retrieval methods.			
Jégou et al. [46] 75.07%	Li et al. [52] 89.20%	Zheng et al. [53] 85.80%	Tolias et al. [54] 82.20%
Qin et al. [55] 84.40%	Zheng et al. [56] 85.20%	**Ranked List Graph Dist.** 85.33%	

4.7. Extension to generic image retrieval tasks

In fact, the Ranked-List Graph Distance and the retrieval model based on ranking information can be used in generic image retrieval tasks. Our technique was designed to be flexible and robust and, hence, the feature input is not limited to any one type. Instead, all possible data types can be used. The only requirement is that the dissimilarity between features must be numerically represented by an appropriate distance metric.

An experiment was conducted for evaluating the effectiveness of the proposed method in generic image retrieval tasks. The Holidays [46] dataset, a popular image retrieval benchmark was considered. The dataset is composed of 1,491 personal holiday pictures and defines 500 queries. The MAP scores are used as effectiveness measures.

Five different image features are considered, including two color descriptors: Auto Color Correlogram (ACC) [47] and Scalable Color Descriptor (SCD) [48]; one color/texture descriptor: Joint Composite Descriptor (JCD) [49]; and two Convolutional Neural Network (CNN) features: Caffe [50] and OverFeat [51].

Table 13 presents the results for the Ranked List Graph Distance. Positive gains can be observed for all considered features, reaching +10.44%. A comparison with state-of-the-art approaches is presented in Table 14, considering the best retrieval results of each approach. The proposed approach method also achieves very high effectiveness scores, comparable or superior to the state-of-the-art.

5. Conclusions

Re-ranking algorithms have been studied a lot recently with the objective of improving the effectiveness of content-based image retrieval tasks. In special for shape retrieval, several research groups have been validating approaches that learn iteratively the similarity/distance among shape objects. One important class of methods relies on the use of graphs and its combination with diffusion approaches for learning the similarity among shapes. Those methods have been demonstrated to be very effective, however, at the same time, computationally costly.

In this paper we addressed this issue, by introducing a novel graph-based model that combines cross-references among shapes in different ranked lists. A single collection graph is defined in terms of the combination of the different available rank-list graphs and then later is used to redefine the distance among shape objects. This process is repeated along iterations.

The efficiency of the method relies on the use of only top-ranked shapes in the ranked lists. The effectiveness, in turn, was demonstrated by the performance of an extensive experimental protocol considering widely used shape collections. Effectiveness experimental results demonstrated that the ranked-list graph model is able to yield significant results when compared with the use of shape descriptors in isolation, being comparable or superior than several state-of-the-art approaches.

Future work will be focused on the application of the proposed method in other searching scenarios involving multimodal information (e.g., multimedia geocoding tasks [57]) or other types of data (e.g., video [58]). We also plan to investigate the use of the ranked-list graph model integrated with indexing schemes [59] to speed up the identification of top-k neighbors.

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] X. Yang, S. Koknar-Tezel, L.J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2009), pp. 357–364.
[2] Z. Tu, S. Zheng, A. Yuille, Shape matching and registration by data-driven EM, Comput. Vis. Image Underst. 109 (2008) 290–304.
[3] R.d. S. Torres, A.X Falcão, Contour salience descriptors for effective image retrieval and analysis, Image Vis. Comput. 25 (2007) 3–13.
[4] G. McNeill, S. Vijayakumar, Hierarchical procrustes matching for shape retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2006), pp. 885–894.
[5] Perceptually motivated shape context which uses shape interiors, Pattern Recognit. 46 (2013) 2092–2102.
[6] Y. Mingqiang, K.K. Idiyo, R. Joseph, A survey of shape feature extraction techniques, Pattern Recognit. (2008) 43–90.
[7] M.S. Drew, T.K. Lee, A. Rova, Shape retrieval with eigen-css search, Image Vis. Comput. 27 (2009) 748–755.
[8] R.d. S. Torres, A.X. Falcão, Contour salience descriptors for effective image retrieval and analysis, Image Vis. Comput. 25 (2007) 3–13.
[9] N. Arica, F.T.Y. Vural, BAS: a perceptual shape descriptor based on the beam angle statistics, Pattern Recognit. Lett. 24 (2003) 1627–1639.
[10] I.E. Rube, N. Alajlan, M. Kamel, M. Ahmed, G. Freeman, Robust multiscale triangle-area representation for 2d shapes, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), vol. 1, 2005, pp. I–545–8.
[11] A. Erdem, S. Tari, A similarity-based approach for shape classification using aslan skeletons, Pattern Recognit. Lett. 31 (2010) 2024–2032.
[12] X. Bai, W. Liu, Z. Tu, Integrating contour and skeleton for shape classification, in: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2005, pp. 360–367.
[13] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 286–299.
[14] X. Bai, X. Yang, L.J. Latecki, W. Liu, Z. Tu, Learning context-sensitive shape similarity by graph transduction, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 861–874.
[15] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, N. Tang, Learning context-sensitive similarity by shortest path propagation, Pattern Recognit. 44 (2011) 2367–2374.
[16] X. Yang, L. Prasad, L. Latecki, Affinity learning with diffusion on tensor product graph, IEEE Trans.Pattern Anal. Mach. Intell., 35 (2013) 28–38.
[17] J. Jiang, B. Wang, Z. Tu, Unsupervised metric learning by self-smoothing operator, in: Proceedings of the International Conference on Computer Vision (ICCV, 2011), pp. 794–801.
[18] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR', 2013), pp. 1320–1327.
[19] B. Wang, J. Jiang, WeiWang, Z.-H. Zhou, Z. Tu, Unsupervised metric fusion by cross diffusion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR', 2012), pp. 3013–3020.
[20] X. Bai, B. Wang, C. Yao, W. Liu, Z. Tu, Co-transduction for shape retrieval, IEEE Trans. Image Process. 21 (2012) 2747–2757.
[21] D.C.G. Pedronette, J. Almeida, R.d. S. Torres, A scalable re-ranking method for content-based image retrieval, Inf. Sci. 265 (2014) 91–104.
[22] X. Bai, S. Bai, X. Wang, Beyond diffusion process: Neighbor set similarity for fast re-ranking, Inf. Sci. 325 (2015) 342–354.
[23] D.C.G. Pedronette, O.A. Penatti, R.d. S. Torres, Unsupervised manifold learning using reciprocal KNN graphs in image re-ranking and rank aggregation tasks, Image Vis. Comput. 32 (2014) 120–130.
[24] D.C.G. Pedronette, R.d. S. Torres, Image re-ranking and rank aggregation based on similarity of ranked lists, Pattern Recognit. 46 (2013) 2350–2360.
[25] Y. Chen, X. Li, A. Dick, R. Hill, Ranking consistency for image matching and object retrieval, Pattern Recognit. 47 (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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR', 2011), pp. 777–784.
[27] S. Zhang, M. Yang, T. Cour, K. Yu, D. Metaxas, Query specific rank fusion for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 803–815.
[28] W. Webber, A. Moffat, J. Zobel, A similarity measure for indefinite rankings, ACM Trans. Inf. Syst. 28 (2010) 20:1–20:38.
[29] R.d. S. Torres, A.X. Falcão, Content-based image retrieval: Theory and applications, Revista de Informática Teórica e Aplicada 13 (2006) 161–185.
[30] H. Jegou, C. Schmid, H. Harzallah, J. Verbeek, Accurate image search using the contextual dissimilarity measure, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 2–11.
[31] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, Q. Tian, Query-adaptive late fusion for image search and person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR', 2015, pp. 1741–1750.
[32] D.C.G. Pedronette, R.d. S. Torres, Exploiting pairwise recommendation and clustering strategies for image re-ranking, Inf. Sci. 207 (2012) 19–34.
[33] D.C.G. Pedronette, R.d. S. Torres, Shape retrieval using contour features and distance optimization, in: Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP, 2010), pp. 197–202.
[34] R. Gopalan, P. Turaga, R. Chellappa, Articulation-invariant representation of non-planar shapes, in: Proceedings of the 11th European Conference on Computer Vision (ECCV, vol. 3, 2010), pp. 286–299.
[35] L.J. Latecki, R. Lakmper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2000), pp. 424–429.
[36] H. Ling, X. Yang, L.J. Latecki, Balancing deformability and discriminability for shape matching, in: Proceedings of the European Conference on Computer Vision (ECCV, vol. 3, 2010), pp. 411–424.
[37] E. Persoon, K. Fu, Shape discrimination using fourier descriptors, IEEE Trans. Syst. Man Cybernet. 7 (1977) 170–179.
[38] F. Mokhtarian, S. Abbasi, Shape similarity retrieval under affine transforms, Pattern Recognit. 35 (2002) 31–41.
[39] F.A. Andaló, P.A.V. Miranda, R.d. S. Torres, A.X. Falcão, Shape feature extraction and description based on tensor scale, Pattern Recognit. 43 (2010) 26–36.
[40] N. Alajlan, I.E. Rube, M.S. Kamel, G.H. Freeman, Shape retrieval using triangle-area representation and dynamic space warping, Pattern Recognit. 40 (2007) 1911–1920.
[41] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, in: Proceedings of the IEEE: International Conference on Computer Vision and Pattern Recognition (CVPR), 2003, pp. 409–415.
[42] Z. Tu, A.L. Yuille, Shape matching and recognition - using generative models and informative features, in: Proceedings of the European Conference on Computer Vision (ECCV, 2004), pp. 195–209.
[43] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 509–522.
[44] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: Proceedings of the Asian Conference on Computer Vision (ACCV, 2009), pp. 655–666.
[45] G.V. Cormack, C.L.A. Clarke, S. Buettcher, 2009. Reciprocal rank fusion outperforms condorcet and individual rank learning methods, in: Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 758–759.
[46] H. Jegou, M. Douze, C. Schmid, Hamming embedding and weak geometric consistency for large scale image search, in: Proceedings of the European Conference on Computer Vision, ECCV, 2008, pp. 304–317.
[47] J. Huang, S.R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlograms, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997, pp. 762–768.
[48] B. Manjunath, J.-R. Ohm, V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Syst. Video Technol. 11 (2001) 703–715.
[49] K. Zagoris, S. Chatzichristofis, N. Papamarkos, Y. Boutalis, Automatic image annotation and retrieval using the joint composite descriptor, in: Proceedings of the 14th Panhellenic Conference on Informatics (PCI), pp. 143–147.
[50] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S.Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding 2014 arXiv preprint arXiv:1408.5093.

[51] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2014, pp. 512–519.
[52] X. Li, M. Larson, A. Hanjalic, Pairwise geometric matching for large-scale object retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR', 2015), pp. 5153–5161.
[53] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: Coupled multi-index for accurate image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2014), pp. 1947–1954.
[54] G. Tolias, Y. Avrithis, H.J.A. gou, To aggregate or not to aggregate: Selective match kernels for image search, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV, 2013), pp. 1401–1408.
[55] D. Qin, C. Wengert, L.V. Gool, Query adaptive similarity for large scale object retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2013), pp. 1610–1617.
[56] L. Zheng, S. Wang, Q. Tian, Coupled binary embedding for large-scale image retrieval, IEEE Trans. on Image Process. 23 (2014) 3368–3380.
[57] L.T. Li, D.C.G. Pedronette, J. Almeida, O.A.B. Penatti, R.T. Calumby, R.S. Torres, A rank aggregation framework for video multimodal geocoding, Multimed. Tools Appl. 73 (2014) 1323–1359.
[58] J. Almeida, D.C.G. Pedronette, O.A.B. Penatti, Unsupervised manifold learning for video genre retrieval, in: Proceedings of the Iberoamerican Congress on Pattern Recognition (CIARP), 2014, pp. 604–612.
[59] J. Almeida, R.S. Torres, N.J. Leite, BP-tree: An efficient index for similarity search in high-dimensional metric spaces, in: Proceedings of the ACM Int. Conf. Information and Knowledge Management (CIKM), 2010, pp. 1365–1368.