A Correlation Graph Approach for Unsupervised Manifold Learning in Image Retrieval Tasks

Daniel Carlos Guimarães Pedronette[a], Ricardo da S. Torres[b]

[a]Department of Statistics, Applied Mathematics and Computing (DEMAC),
São Paulo State University (UNESP), Rio Claro, Brazil
[b]RECOD Lab, Institute of Computing (IC),
University of Campinas (UNICAMP), Campinas, Brazil

Abstract

Effectively measuring the similarity among images is a challenging problem in image retrieval tasks due to the difficulty of considering the dataset manifold. This paper presents an unsupervised manifold learning algorithm that takes into account the intrinsic dataset geometry for defining a more effective distance among images. The dataset structure is modeled in terms of a Correlation Graph (CG) and analyzed using Strongly Connected Components (SCCs). While the Correlation Graph adjacency provides a precise but strict similarity relationship, the Strongly Connected Components analysis expands these relationships considering the dataset geometry. A large and rigorous experimental evaluation protocol was conducted for different image retrieval tasks. The experiments were conducted in different datasets involving various image descriptors. Results demonstrate that the manifold learning algorithm can significantly improve the effectiveness of image retrieval systems. The presented approach yields better results in terms of effectiveness than various methods recently proposed in the literature.

Keywords: content-based image retrieval, unsupervised manifold learning, correlation graph, strongly connected components

1. Introduction

A huge amount of visual content has been accumulated daily, generated from a large variety of digital sources [1, 2], from personal mobile cameras to interactive video-games and surveillance devices. In this scenario, the demand for methods capable of understanding the image and video content is increasing. Human action recognition [3] and pose estimation [4] methods, for example, have achieved relevant advances in human-computer interaction applications and human behavior analysis.

Regarding image collections, a change of behavior can be observed, since common users are not long mere consumers and have become active producers of digital content, especially images. The Content-Based Image Retrieval (CBIR) systems are considered a promising solution, supporting searches capable of taking into account the visual properties of digital images [2, 5]. The main objective of these systems is to retrieve relevant collection images ranked according to their similarity to a query input (e.g., query image).

For decades, the development of CBIR systems have been mainly supported by the creation of various visual features (based on shape, color, and texture properties) and different distance measures [5]. A major challenge in this scenario, known as *semantic gap*, consists in the difficulties in mapping low-level features to high-level concepts typically found within images. The semantic gap [6] affects a broad class of applications, from image retrieval tasks to complex event detection in video sequences. In these scenarios, it is beneficial to consider the common knowledge from both the low-level features and the high-level concept information [7]. More recently, aiming at improving the retrieval effectiveness by reducing the semantic gap, research initiatives have focused on other stages of the retrieval process, which are not directly related to low-level feature extraction procedures [8].

In several computer vision and image retrieval applications, images are represented by feature vectors and modeled as high dimensional points in an Euclidean space. Handling high dimensional feature representations consists in a challenging task. Common approaches aim at both reducing the complexity of a data set and preserving information that is important for understanding the data structure itself [9]. Dimensionality reduction methods, for example, aims at finding meaningful low-dimensional structures hidden in their high-dimensional observations [10, 11].

For images represented in high dimensional spaces, their comparison is of-

2

ten based on the use of distance functions applied on their corresponding feature vectors. In retrieval scenarios, for example, images are commonly ranked in increasing order of their distances to defined query patterns. Distance functions usually consider pairs of images. However, the pairwise distance analysis is very simplistic, since it provides only locally restrict comparisons and ignores more global relationships and the dataset structure itself.

In fact, since collection images are often encoded in a much lower-dimensional intrinsic space, capturing and exploiting the intrinsic manifold structure becomes a central problem for different vision, learning, and retrieval tasks [12, 13, 14]. In this scenario, methods have been proposed with the aim of ranking collection objects with respect to the intrinsic global manifold structure [14]. Diverse methods also have been proposed in order to improve the effectiveness of distance measures in image retrieval tasks [15, 16, 17, 12]. In general, these approaches aims at replacing strategies based on pairwise distance computations by more global affinity measures capable of considering the dataset manifold [17].

In practical terms, the unlabeled information encoded in the dataset manifold can be exploited for reducing the effects of the semantic gap. However, while user preferences [18] and relevance feedback information [19] are commonly used for reducing the semantic gap, the use of unlabeled information is much more challenging [20].

In this paper, we discuss a novel unsupervised manifold learning algorithm, which aims at miming the human behavior in judging similarity. The unconscious mind captures incomplete data from senses and completes the missing information based on the context [21, 22, 23]. Analogously, in the retrieval scenario image descriptors often provide incomplete similarity information. The proposed algorithm exploits unlabeled contextual information encoded in the dataset manifold through the Correlation Graph for improving the effectiveness of distance/similarity measures. In this sense, the context can be seen as any complementary information about similarity among images, as the set of images in a strongly connected component.

The algorithm is based on the Correlation Graph (CG) and Strongly Connected Components (SCCs) and takes into account the intrinsic geometry of the dataset manifold by computing a Correlation Graph Distance. A graph representation of the dataset is constructed by exploiting the correlation information among images, which provides strong indicators of similarity. Correlation indices are computed by analyzing the distances and top positions of ranked lists. Strongly Connected Components are computed based on the

3

Correlation Graph with the aim of expanding the similarity neighborhood and discovering the intrinsic geometry of the dataset manifold. To the best of our knowledge, this is the first method for unsupervised manifold learning using correlation graphs and strongly connected components.

The low computational efforts required by the proposed method also represent a significant advantage when compared to related work [17, 15, 24, 25, 26, 27]. Diffusion-based approaches [28, 27, 15, 17] require the computation of powers of the transition matrix or matrix inversion procedures. Graph-based methods [26] compute the shortest paths independently of each query element increasing the computational costs. Iterative re-ranking methods, in turn, [24, 25, 29] require successively sorting steps. Unlike other methods, our approach computes a new distance among images considering only different correlation thresholds, without the need of performing successively distance computing procedures. In addition, the proposed method considers only the top-ranked images, which represent a smaller number of elements when compared with the number of objects handled in recently proposed methods [24]. The low computational costs required is mainly due to the strongly connected components analysis, through which it is possible considering geometry of the entire dataset manifold by using a small neighborhood set.

This paper differs from our previous work [30] as it presents a deeper and broader analysis of the unsupervised manifold learning algorithm. In addition to the distance correlation analysis [30] using the traditional Pearson correlation coefficient, we also exploit recently proposed rank correlation measures [31] for constructing the Correlation Graph. In this way, the proposed method can also be modeled using only ranking information, without the need of distance scores. A normalization of similarity scores is included and it is demonstrated that this step improves the effectiveness results. In addition, a distance fusion approach is also proposed and the experimental evaluation is updated reflecting the novel contributions and expanded for considering other datasets.

An extensive experimental evaluation was conducted, considering six public datasets and 22 image descriptors, including global shape, color and texture descriptors, local descriptors, and convolution-neural-network-based descriptors. Experiments were conducted on different retrieval tasks involving general retrieval, object retrieval, and multimodal retrieval tasks, considering visual and textual descriptors. We also evaluated the proposed manifold learning method in comparison with various other state-of-the-art ap-

proaches. Results from a rigorous experimental protocol show that the proposed method can achieve significant effectiveness gains (up to +38%) yielding better results in terms of effectiveness performance than state-of-the-art approaches.

The paper is organized as follows: In Section 1, we present the proposed unsupervised manifold learning method. Section 3 presents the experimental evaluation and, finally, Section 4 discusses the conclusions and presents future work.

2. Manifold Learning By Correlation Graph

The main objective of the discussed unsupervised learning algorithm is to model the intrinsic geometry of a dataset manifold in terms of a correlation analysis and its correspondent connectivity. The use of the Correlation Graph and Strongly Connected Components in manifold learning tasks was first introduced by the authors in a conference paper [30]. However, the Correlation Graph was originally constructed based on the correlation among distances. In this work, a more comprehensive model is proposed, allowing the use of different correlation indices, including rank-based measures. The novel proposed method also allows the combination of different features in distance fusion tasks. A new effectiveness estimation measure is proposed in order to give higher weights to more effective features in fusion tasks.

2.1. Overview

An overview of the proposed graph-based approach is presented in this section. The method can be roughly divided into four steps:

1. The correlation between each dataset image and the images placed at top positions of its ranked list is computed. The method is able to consider both distance and rank correlation measures. For each ranked list, only a small set of the most correlated images, which are the most likely to be similar to the query image, are selected based on a correlation threshold. The adjacency of the *Correlation Graph* (CG) is defined according to the selected images.

2. The Correlation Graph is analyzed for identifying *Strongly Connected Components* (SCCs). The use of SCCs aims at expanding the neighborhood set of similar images by taking into account the intrinsic geometry of the dataset manifold.

5

3. The first two steps are repeated using different values of correlation thresholds. For each value, a similarity score is incrementally computed by combining information from the correlation graph adjacency and SCCs. The increments are defined according to the confidence defined by the thresholds.

4. A new distance, named *Correlation Graph Distance*, is computed by exploiting the similarity scores computed for different correlation thresholds.

While the edges defined by the Correlation Graph provide a very strong indication of similarity (specially for high correlation thresholds), the edges include a very small neighborhood set. The SCCs allow for the expansion of the neighborhood considering the dataset structure. The capacity of the proposed method of considering the geometry of the dataset manifold is illustrated in Figures 1, 2, and 3.

Figure 1 illustrates the Two-Moon dataset considering the Euclidean distance. One point is selected as a labeled point in each moon (marked with a filled circle and triangle). In the following, all other data points are assigned to the closest labeled point, determining their neighborhood. As it can be observed, the extremities of the moons are misclassified, since the Euclidean distance does not consider the geometry structure of the dataset. Figure 2 illustrates an intermediary step of the proposed method. Points with edges to the labeled point in the Correlation Graph are marked with stars, the SCCs are illustrated in colors (blue and red) and the unclassified points are illustrated in green. Figure 3 illustrates the final configuration that considers the distances computed using the *Correlation Graph Distance*. We can observe that the ideal classification, which respects the whole geometry of the dataset manifold, was produced.

2.2. Correlation Graph

This section formally describes the construction of the Correlation Graph (CG). Let $\mathcal{C}=\{img_1,\, img_2,\, \ldots,\, img_n\}$ be an image collection, where n is the size of the collection.

Let $\mathcal{D}$ be an image descriptor which can be defined [32] as a tuple $\mathcal{D} = (\epsilon, \rho)$, where $\epsilon\colon \hat{I} \to \mathbb{R}^n$ is a function, which extracts a feature vector $v_{\hat{I}}$ from an image $\hat{I}$; and $\rho\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a distance function that computes the distance between two images according to the distance between their
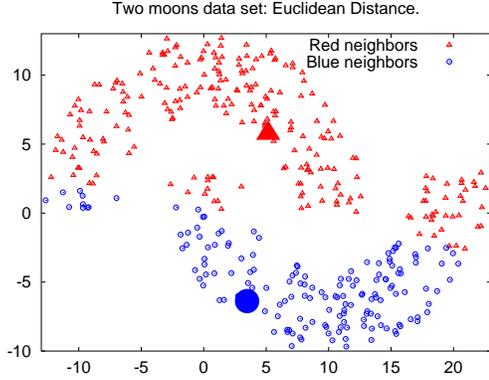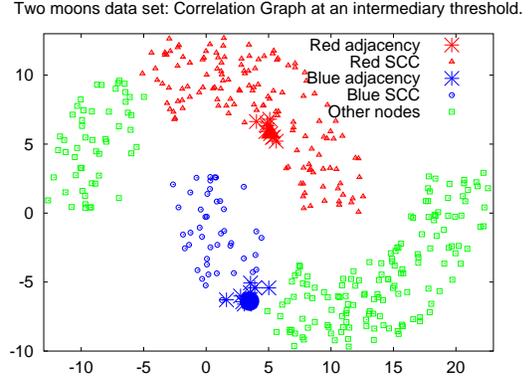
Figure 1: Euclidean distance.
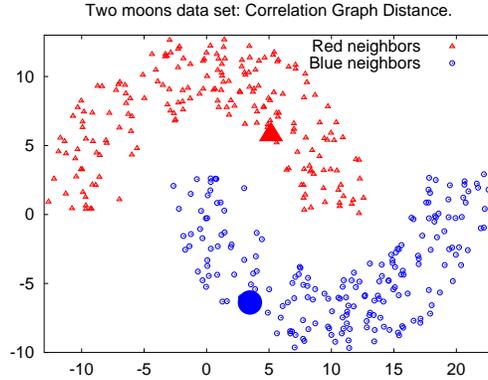

Figure 2: Intermediary CG structures.

Figure 3: Correlation Graph Distance.

corresponding feature vectors. For readability purposes, the notation $\rho(i, j)$ is used to refer to the distance between two images img_i and img_j along the paper.

Let $\tau_q = (img_1, img_2, \ldots, img_{n_s})$ be a ranked list computed based on the distance function ρ, in response to a query image img_q. The ranked list τ_q can be defined as a permutation of the subset $\mathcal{C}_s \subset \mathcal{C}$, where the subset $\mathcal{C}_s$ contains the n_s most similar images to query image img_q, and therefore, $|\mathcal{C}_s| = n_s$ and $n_s \ll n$. We interpret $\tau_q(i)$ as the position (or rank) of image img_i in the ranked list τ_q.

The *Correlation Graph* (CG) is a directed graph $G = (V, E)$, where the

set of vertices V is defined by the image collection $\mathcal{C}$, such that each image is represented by a node and $V = \mathcal{C}$. The edge set E is defined considering the correlation among images at the top n_s positions of each ranked list, as follows:

$$E = \{(img_q, img_j) \mid \tau_q(j) \leqslant n_s \wedge cor(q, j) \geqslant t_c\}, \tag{1}$$

where $cor(q, j)$ defines a correlation measure between img_q and img_j and t_c is the correlation threshold considered. Therefore, there will be an edge from img_q to img_j, if: (i) img_j is at the top-n_s positions of ranked of img_q; and (ii) the correlation score between them are greater than a given threshold t_c. Notice that, the smaller the correlation threshold is, the denser the constructed graph is. That increases the neighborhood adjacency but also increases the number of false positives. On the other hand, the higher the threshold becomes, the sparser is the graph. That, however, leads to a higher confidence.

The proposed manifold learning algorithm is robust to the use of different correlation measures for computing the $cor(q, j)$ coefficient. Next subsections detail two different distance and rank correlation approaches.

2.2.1. Distance Correlation

Given two similar images img_q, img_j, the distances from these images to other similar images in common is frequently low. In other words, we can say that the distances to their neighborhood is correlated. In our approach, the distance correlation between them is measured using the Pearson's Correlation Coefficient and considering the distances to the k-nearest neighbors of img_q and img_j.

Let $\mathcal{N}_k(q)$ be the set containing the k-nearest neighbors to given image img_q. Let $\mathcal{N}_k(q, j)$ be the union set containing the k-nearest neighbors of both images img_q and img_j, such that $\mathcal{N}_k(q, j) = \mathcal{N}_k(q) \cup \mathcal{N}_k(j)$. We define two vectors X and Y containing, respectively, the distances from images img_q and img_j to each image $img_i \in \mathcal{N}_k(q, j)$. Let img_i be the i-th image of the set $\mathcal{N}_k(q, j)$, we define $X_i = \rho(q, i)$ and $Y_i = \rho(j, i)$. The correlation $cor(q, j)$ score is defined as follows:

$$cor(q, j) = \frac{\sum_{i=1}^{k_u}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{k_u}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{k_u}(Y_i - \overline{Y})^2}}. \tag{2}$$

8

The value of the Pearson's correlation coefficient ranges from -1 to $+1$, where $+1$ indicates a perfect positive linear relationship. We normalize the computed value of $cor(q,j)$ in the interval $[0,1]$. Figure 4 illustrates the distance correlation analysis using the Pearson Correlation Coefficient for computing the CG adjacency.
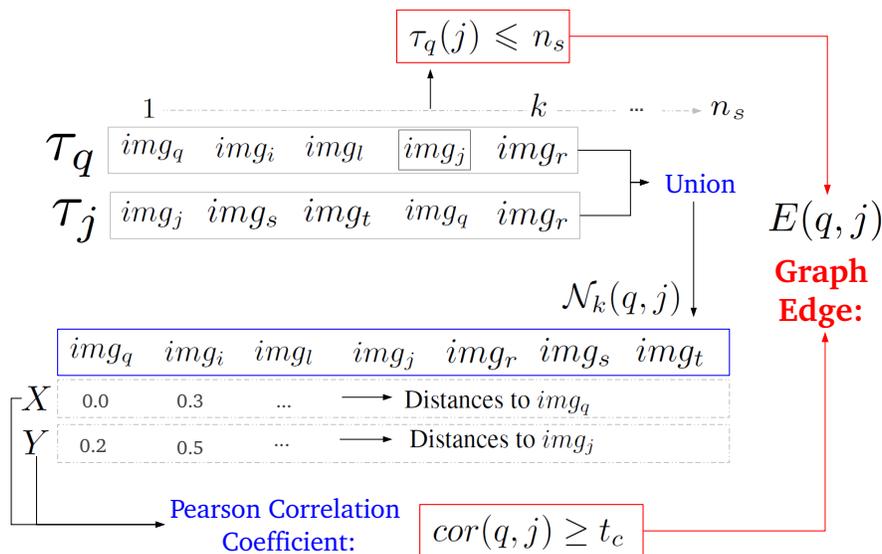

Figure 4: Distance correlation analysis for computing the adjacency of the Correlation Graph.

2.2.2. Rank Correlation

Ranked lists, which are commonly used for visualizing retrieval results, can also be exploited for correlation analysis. Ranked lists encode relevant contextual information [24] by defining relationships among several images, and not only pairs of images as distance functions. In addition, an advantage of using ranking information refers to the fact that there is no need of computing distance scores, which usually may be computed in different ranges, requiring therefore normalization procedures.

A recently proposed measure [31] based on a simple probabilistic user model is used for performing the rank correlation analysis. The Rank-Biased Overlap [31] (RBO) compares the overlap of two rankings at incrementally

increasing depths. This measure takes a parameter that specifies the probability of considering the overlap at the next level. The weight of the overlap measured at each depth is computed based on these probabilities. The correlation score according to the RBO measure is defined as follows:

$$cor(q, j) = (1 - p) \sum_{d=1}^{k} p^{d-1} \times \frac{|\mathcal{N}_k(q) \cap \mathcal{N}_k(j)|}{d}, \tag{3}$$

where p is a constant, which determines the strength of the weighting to top ranks.

2.3. Strongly Connected Components

The edges of the Correlation Graph represent a very strong indication of similarity among images for high thresholds. However, although very precise, the edges include a very small neighborhood set. We can observe this behavior in Figure 2. In this scenario, for ensuring the comprehensiveness of the retrieval process is necessary to expand the similarity neighborhood. Strongly Connected Components (SCCs) are considered for this task as it encodes the geometry of the dataset manifold. Recently, the reciprocal neighborhood [33, 34] has been exploited in image retrieval tasks for analyzing the dataset structure through reciprocal references. We use the SCCs of the Correlation Graph with an analogous objective, since SCCs also define a reciprocal connectivity relationship among a set of nodes.

The strongly connected components of a directed graph are defined by subgraphs that are themselves strongly connected, i.e., where every vertex is reachable from every other vertex. Formally, let S be a SCC, for every pair img_q, $img_j \in S$ there is an oriented path between img_q and img_j and vice versa.

The Tarjan [35] algorithm, which is linear on the size of the graph, is used for computing the SCCs. Each SCC is defined as a set of images S_i. The overall output of the algorithm is a set of SCCs $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$. The information provided by SCCs is exploited for computing the Correlation Graph Distance.

2.4. Correlation Graph Distance

The objective of the Correlation Graph Distance is to exploit all information encoded in the Correlation Graph and SCCs for computing a new and

10

more effective distance among images. In this way, we define a Correlation Graph Similarity Score, which aims at quantifying the association between images according to the Correlation Graph and SCCs. The similarity score between two given images img_i, img_j is defined by $W_{i,j}$, given by an sparse similarity matrix W.

The similarity score $W_{i,j}$ is computed in terms of increments, according to the adjacency defined by Correlation Graph and correspondent SCCs. The sparsity of the matrix W is given by the fact that increments are computed for non-similar images, and therefore, most of the elements of the affinity matrix remain empty. Let $E(q)$ denote a set of images to whom img_q have edges in the Correlation Graph, the similarity score between img_i, $img_j \in E(q)$ receives an increment, according to the correlation threshold t_c considered. The same increments are computed for images that belong to a same SCC, i.e., for every pair img_q, $img_j \in S$.

The correlation threshold t_c defines the magnitude of the similarity increment since it provides an unsupervised estimation of the confidence of connectivity information of the Correlation Graph.

Algorithm 1 outlines the proposed method for computing the similarity score $W_{i,j}$. Lines 4-11 define the similarity increments according to the CG adjacency, while Lines 13-17 exploit information from SCCs for computing other increments. Different threshold values (t_c) are considered, according to the external loop (Lines 2-19).

Figure 5 illustrates an example of the Correlation Graph and its capacity of discovering new relationships among images. The solid lines represent the graph adjacency, according to the correlation conditions. As previously discussed, these edges provide a strong indication of similarity, and therefore, the images img_i, img_j, img_l are probably similar as well. In this scenario, the increments are illustrated by dashed lines.

Based on computed similarity scores a normalization procedure is defined. Since the Correlation Graph is oriented and the adjacency is not symmetric, the similarity scores may present high variations. Therefore, a normalization step is performed according to Equation 4. The normalized similarity score $W_{N_{i,j}}$ is computed proportionally to the accumulated increments received for each image.

Algorithm 1 Correlation Graph Distance

Require: Correlation Graph $G = (V, E)$, Set of SCCs $\mathcal{S}$
Ensure: Correlation Graph Similarity Score $W_{i,j}$

1:   $t_c \leftarrow t_{start}$
2: **while** $t_c \leqslant 1$ **do**
3:     { Correlation Graph }
4:     **for all** $img_q \in V$ **do**
5:       **for all** $img_i \in E(q)$ **do**
6:         $W_{q,i} \leftarrow W_{q,i} + t_c$
7:         **for all** $img_j \in E(q)$ **do**
8:           $W_{i,j} \leftarrow W_{i,j} + t_c$
9:         **end for**
10:      **end for**
11:    **end for**
12:    { Strongly Connected Components }
13:    **for all** $S_c \in \mathcal{S}$ **do**
14:      **for all** $img_i, img_j \in S_c$ **do**
15:        $W_{i,j} \leftarrow W_{i,j} + t_c$
16:      **end for**
17:    **end for**
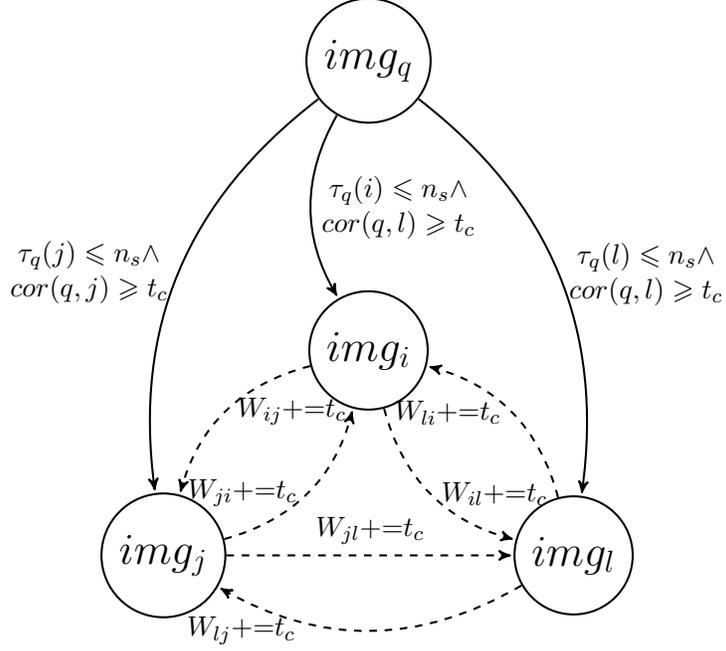18: $t_c \leftarrow t_c + t_{inc}$
19: **end while**

Figure 5: Example of correlation graph and similarity increments.

$$W_{N_{i,j}} = \frac{W_{i,j}}{\sum\limits_{c=1}^{n} W_{j,c}}. \tag{4}$$

Finally, based on the normalized similarity score $W_{N_{i,j}}$, the *Correlation Graph Distance* $\rho_c(i,j)$ is computed as follows:

$$\rho_c(i,j) = \frac{1}{1 + W_{N_{i,j}}}. \tag{5}$$

Notice that the similarity matrix is sparse, since most of the similarity scores $W_{N_{i,j}}$ between images $imag_i, img_j \in \mathcal{C}$ do not receive increments. For these cases, the ranked lists are organized according to the initial ranked lists.

13

2.5. Correlation Graph for Distance Fusion

Different features often encode distinct and complementary visual information extracted from images. Therefore, it is intuitive that, for a given query, if a feature is associated with good effectiveness scores by itself and is complementary (heterogeneous) to other features, then it is expected that a higher search accuracy can be achieved by combining them [36].

In fact, many recent works [36, 37, 29] have been demonstrating the effectiveness of fusion approaches in image retrieval tasks. An important challenge for fusion methods consists in estimating the quality of retrieval results obtained by distinct features [34, 36, 38]. Other relevant aspect consists in performing the distance fusion considering information encoded in the dataset manifold.

In this work, we present a distance fusion approach, which aims at addressing both relevant aspects at the same time. We propose a simple and effective measure for estimating the quality of retrieval results. Our approach is inspired by the observation [36] that good features tend to better differentiate the top positions of ranked lists, producing a higher score variability.

Aiming at measuring the score variability for a given feature, we exploit the relation between the accumulated distances for images at top-k and for images until the n_s positions. More effective features are expected to produce a very low accumulated top-k distances in comparison with the subset contained until the n_s positions. The accumulated distances for k and n_s positions are formally defined in Equations 6 and 7, respectively.

$$\rho_k(q) = \sum_{img_i \in \mathcal{N}_k(q)} \rho(q, i). \tag{6}$$

$$\rho_{n_s}(q) = \sum_{img_i \in \mathcal{N}_{n_s}(q)} \rho(q, i). \tag{7}$$

The estimation measure $e(q)$ for a query image img_q is defined as follows:

$$e(q) = 1 - \frac{\rho_k(q)}{\rho_{n_s}(q)}. \tag{8}$$

In the following, the distances from different features are combined using a late fusion approach. A multiplicative approach inspired on recent positive results [29, 24] is used for combining the distances. The query-adaptive quality estimation $e(q)$ is used as exponent for assigning weights for each feature distance, as follows:

$$\rho_f(q, i) = \prod_{j=1}^{d} \rho(q, i)^{e(q)}.$$ (9)

Finally, once the different distances have been fused on a single distance (or corresponding ranked lists), the Correlation Graph Distance is performed for considering the relationships encoded in the the dataset manifold.

3. Experimental Evaluation

In this section, we present the results of a large experimental evaluation conducted for assessing the effectiveness of the proposed method. Various different datasets and descriptors are considered aiming at evaluating the algorithm in diverse retrieval tasks. Experiments were conducted on six public image datasets commonly used in the literature, considering 22 different image descriptors. The visual features include global shape, color, and texture features. Searching scenarios considering local descriptors, descriptors based on the use convolutional neural networks, and multimodal retrieval involving visual and textual descriptors are also considered. A rigorous experimental protocol was employed, involving all image datasets and statistical tests. Table 1 summarizes the datasets and descriptors considered.

Figure 6 illustrates sample images from different image datasets used in the experimental evaluation. Each row contains images from a dataset, following the same order of Table 1.

The remainder of this section is organized as follows: Section 3.1 discusses the parameter values and Section 3.2 discusses the impact of the algorithm on distance distribution. Section 3.3 presents the experimental results for

Table 1: Summary of datasets used on the experimental evaluation.

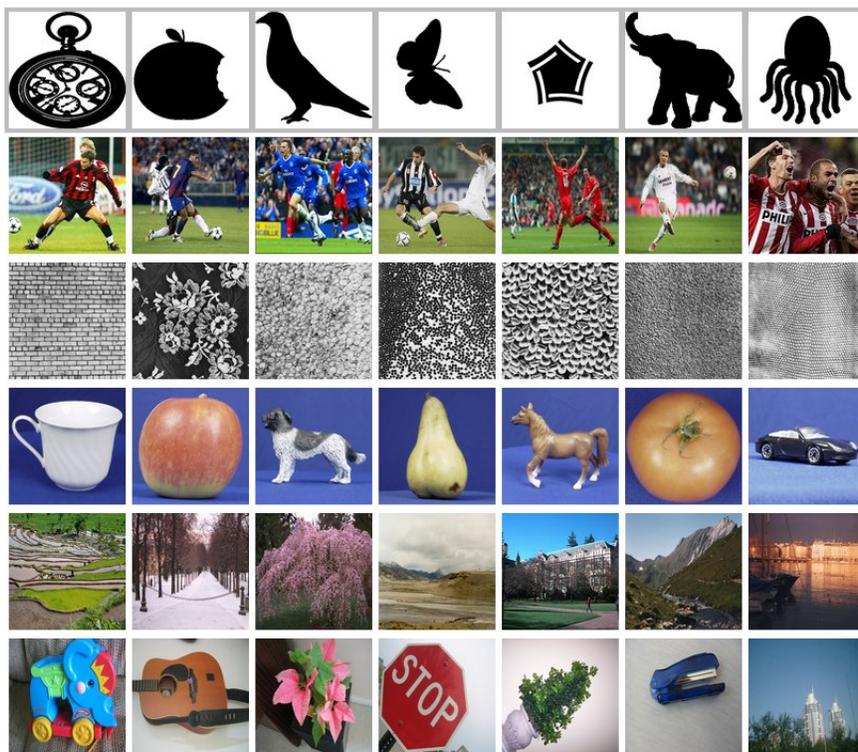| Dataset	Type	Size	Descriptors
MPEG-7 [39]	Shape	1,400	6
Soccer [40]	Color	280	3
Brodatz [41]	Texture	1,776	3
ETH-80 [42]	Color Objects	3,280	4
UW Dataset [43]	Color Scenes + Keywords	1,109	6 visual; 6 textual
UKBench [44]	Color Objects/Scenes	10,200	7

Figure 6: Sample images from various datasets considered.

the proposed approach considering various shape, color, and texture descriptors. Section 3.4 presents the experimental results for object retrieval tasks. Section 3.5 presents results for natural image retrieval, while Section 3.6 discusses the results for multimodal retrieval tasks. Finally, Section 3.7 presents a comparison of the proposed approach with state-of-the-art related methods.

3.1. Impact of Parameters

This section aims at evaluating the robustness of the method to different parameter settings and defining the best parameter values. We conducted various experiments considering the MPEG-7 [39] shape dataset (described in Section 3.3). The first experiment evaluates the impact of the parameters k (size of the neighborhood set used for correlation analysis) and t_{start} (start value of correlation threshold t_c).
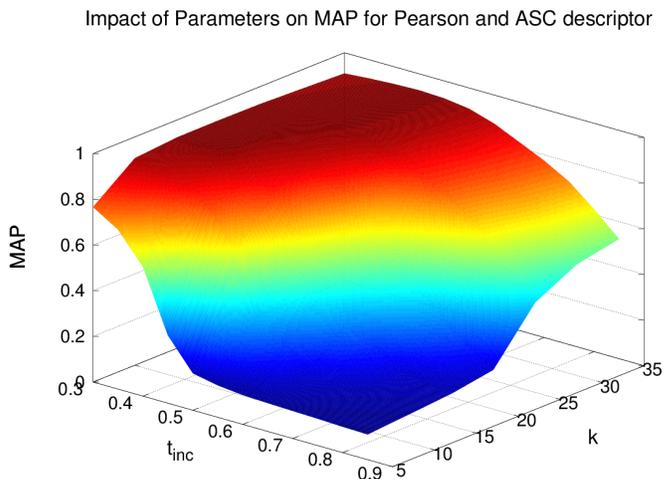

Figure 7: Impact of parameters k and correlation threshold.

Figure 7 illustrates the variations of effectiveness scores given by Mean Average Precision (MAP) according to variations of k and t_{start}. The experiment considered the Aspect Shape Context (ASC) [45] shape descriptor and the Pearson correlation coefficient. We can observe a large red region indicating high retrieval scores for different parameter settings, which demonstrates the robustness of the proposed method. An analogous experiment was conducted considering the RBO rank correlation measure with similar results. In most of remaining experiments, we used the values of $t_{start} = 0.35$ and $t_{start} = 0.05$ for the Pearson and RBO measures, respectively. For the size of the neighborhood, we used $k = 25$ for both measures.

We also evaluate the impact of the size of ranked lists (n_s) and the threshold increments (t_{inc}) on effectiveness gains, considering two shape descriptors: Aspect Shape Context (ASC) [45] and Articulation-Invariant Representation (AIR) [46]. Figures 8 and 9 illustrates the impact of these parameters on the

17

MAP scores. As we can observe, only a small subset of ranked lists is enough to achieve high effectiveness results. The value $n_s = 200$ is used in all other experiments.
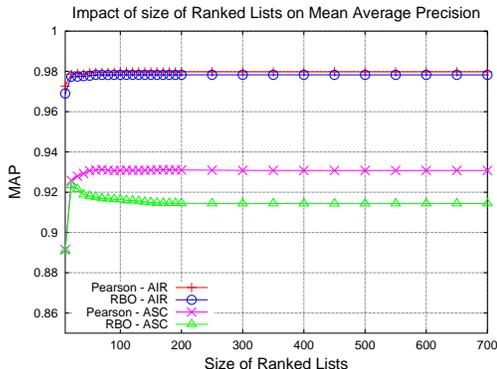

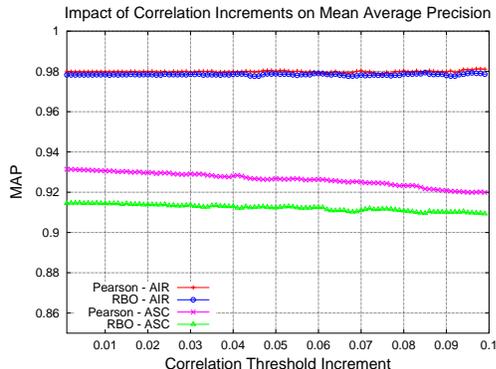Figure 8: Impact of the size of ranked lists (n_s) on effectiveness results.

Figure 9: Impact of correlation increments (t_{inc}) on effectiveness results.

The correlation increment t_{inc} defines the granularity of correlation analysis, representing a smooth trade-off between efficiency and effectiveness. For large increments, the algorithm executes faster, but its effectiveness tends to decrease. Figure 9 illustrates the variations of effectiveness scores according to different values of correlation increment. Notice that the figure presents a very stable behavior. Different values of t_{inc} cause a very small impact on the effectiveness score. We used $t_{inc} = 0.005$ for all datasets and different descriptors.

3.2. Impact of the Algorithm on Distance Distribution

This section discusses the impact of the proposed manifold algorithm on distance distribution. For analyzing this impact, we present a bi-dimensional representation of the MPEG-7 [39] dataset before and after the execution of the algorithm. We selected two arbitrary images, named as reference images, and represent all collection images in the bi-dimensional space. The position of images into this space 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 we want to represent in the bi-dimensional space, the position (x, y) of img_l is defined as $(\rho(img_i, img_l), \rho(img_j, img_l))$.

Figure 10 (a) illustrates the similar reference images. Figure 10 (b) illustrates the respectively distance distribution, presenting similar images 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. Figure 10 (c) illustrates the distance distribution considering the proposed Correlation Graph Distance. Notice the capacity of the proposed algorithm of considering the dataset manifold, which increases the separability between similar and non-similar images.

Figure 10: Impact of the algorithm on distances distribution for similar reference images: **(a)** Similar Reference Images img_i and img_j (fly-2.gif and fly-3.gif) from the MPEG-7 [39] dataset; **(b)** Original distances distribution; **(c)** Distances distribution after the execution of the proposed algorithm.
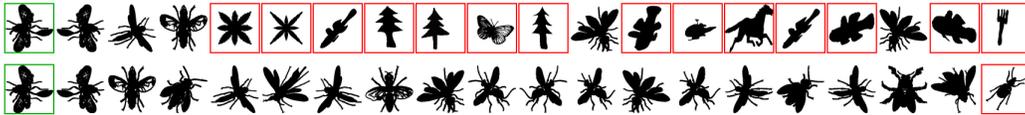
Figure 11: Visual example of the effectiveness gain. Retrieval results before (first row) and after the use of the algorithm (second row). Query image (fly-2.gif) from the MPEG-7 [39] dataset with green border and wrong images with red borders.
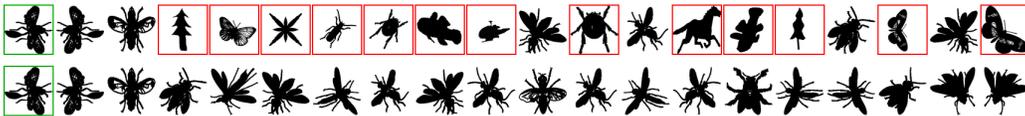
Figure 12: Visual example analogous to Figure 11, considering other query image (fly-3.gif) from the MPEG-7 [39] dataset.

19

The effects of distance distribution on visual retrieval results can be observed in Figures 11 and 12. The figures present the reference images as queries with green borders and the retrieval results with non-similar images with red borders. For each figure, the first and second rows represent the results before and after the use of the algorithm.

The improvements on the effectiveness and the positive effects of the proposed method in the semantic gap problem are remarkable. The original low-level shape features retrieve various different images with several saliences, without any distinction among different meanings. On the contrary, the proposed method successfully exploits the query context and the similarity among most of responses for improving the results and retrieving images from the class "*fly.*"

We also perform this analysis considering two non-similar reference images, which are illustrated in Figure 13 (a). The distance distribution before the use of of the algorithm is illustrated in Figure 13 (b), which presents the similar images to each reference images in red and green circles. Again, we can observe that these images are mixed with non-similar images (in blue). Figure 13 (c) presents the distance distribution considering the Correlation Graph Distance. Analogously to what was observed for similar images, the separability among similar and non-similar images is drastically increased. The impact of the Correlation Graph Distance on the retrieval results can also be observed in Figure 14.

3.3. General Image Retrieval Tasks

This section presents the effectiveness results of the proposed method on general image retrieval tasks, considering shape, color, and texture descriptors. All collection images of each dataset are considered as query images and the Mean Average Precision (MAP) is used as effectiveness measures in most of experiments. We report the relative gains obtained by the use of the proposed manifold learning algorithm for each descriptor and dataset. We also conducted statistical paired t-tests, aiming at assessing the difference between the retrieval results before and after the use of the algorithm is statistical significant.

3.3.1. Shape Retrieval

The MPEG-7 [39] dataset is a well-known shape dataset, composed of 1,400 shapes which are grouped into 70 classes, with 20 objects per class.
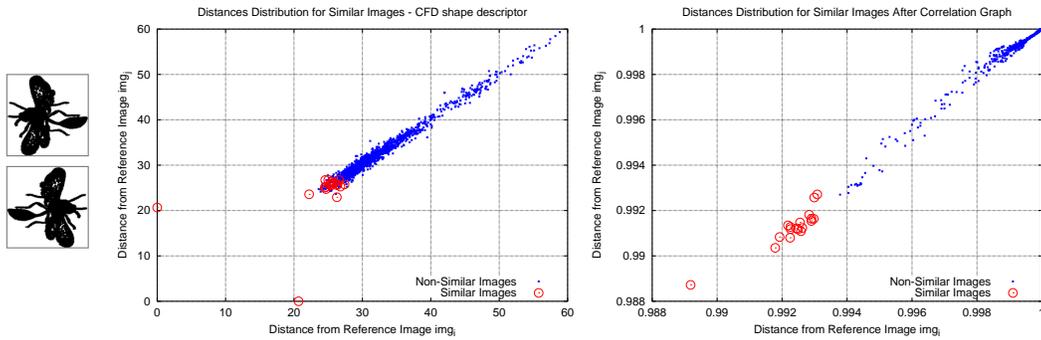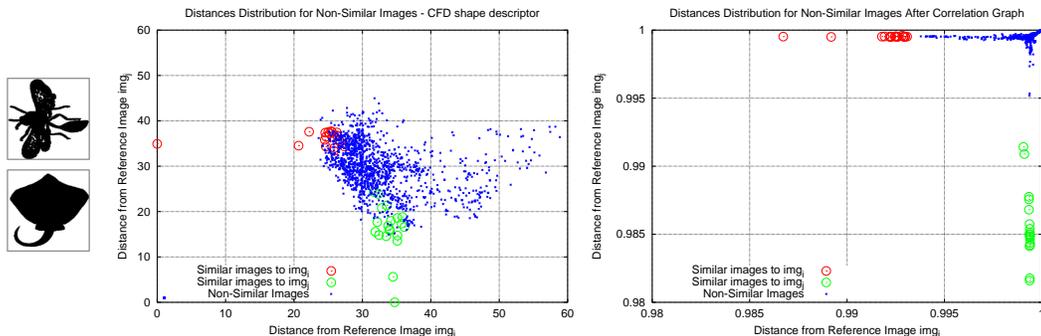
Figure 13: Impact of the algorithm on distances distribution for non-similar images: **(a)** Non-similar reference images img_i and img_j (fly-2.gif and ray-16.gif) from the MPEG-7 [39] dataset; **(b)** Original distances distribution; **(c)** Distances distribution after the execution of the proposed algorithm.
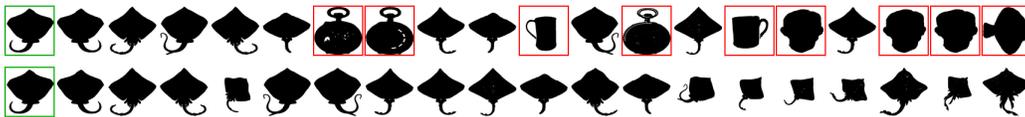
Figure 14: Visual examples of retrieval results before and after the algorithm, considering the query image ray-16.gif.

The dataset is widely used for shape retrieval and post-processing methods evaluation.

The proposed manifold learning algorithm was evaluated on the MPEG-7 [39] dataset, considering six different shape descriptors: Segment Saliences (SS) [47], Beam Angle Statistics (BAS) [48], Inner Distance Shape Context (IDSC) [49], Contour Features Descriptor (CFD) [50], Aspect Shape Context (ASC) [45], and Articulation-Invariant Representation (AIR) [46].

Two effectiveness measures were considered the for the MPEG-7 [39] dataset: the MAP and the bull's eye score, commonly used for this dataset. This 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. Notice that the bull's eye score is equivalent to recall at 40.

Table 2 presents the results considering the bull's eye score of evaluated descriptors. Also positive for all descriptors, the gains range from +6.76% to +33.01% for the Pearson, and from +6.70% to +29.51% for the RBO measure. The paired t-tests indicated statistical significance at 99% for the

21

Table 2: Correlation Graph Distance on the MPEG-7 [39] dataset, considering the Bull's Eye Score (Recall@40).

Shape Descriptor	Bull's Eye Score (Recall@40)	Correlation Graph Dist. Pearson	Gain	Statistic Signific. 99%	Correlation Graph Dist. RBO	Gain	Statistic Signific. 99%
SS [47]	43.99%	58.51%	+33.01%	●	56.97%	+29.51%	●
BAS [48]	75.20%	87.85%	+16.82%	●	84.29%	+12.09%	●
IDSC [49]	85.40%	92.49%	+8.30%	●	92.04%	+7.78%	●
CFD [50]	84.43%	94.84%	+12.33%	●	94.00%	+11.33%	●
ASC [45]	88.39%	95.50%	+8.04%	●	94.47%	+6.88%	●
AIR [46]	93.67%	100%	+6.76%	●	99.95%	+6.70%	●

results of all experiments.

Table 3 presents the MAP scores for both distances an rank correlation measures (Pearson Correlation Coefficient and RBO, respectively). Significant positive gains are observed for all descriptors, ranging from +9.17% to **+38.23%** for the Pearson, and from +7.23% to +35.70% for the RBO measure.

Table 3: Correlation Graph Distance for general image retrieval tasks, considering shape, color, and texture descriptors. Mean Average Precision (MAP) as effectiveness measure and t-tests for evaluating statistical significance.

Descriptor	Dataset	Original Score (MAP)	Correlation Graph Dist. Pearson	Gain	Statistic Signific. 99%	Correlation Graph Dist. RBO	Gain	Statistic Signific. 99%
Shape Descriptors								
SS [47]	MPEG-7	37.67%	52.07%	+38.23%	●	51.12%	+35.70%	●
BAS [48]	MPEG-7	71.52%	83.25%	+16.40%	●	80.18%	+16.40%	●
IDSC [49]	MPEG-7	81.70%	90.10%	+10.28%	●	89.31%	+9.31%	●
CFD [50]	MPEG-7	80.71%	92.52%	+14.63%	●	91.77%	+13.70%	●
ASC [45]	MPEG-7	85.28%	93.10%	+9.17%	●	91.45%	+7.23%	●
AIR [46]	MPEG-7	89.39%	97.98%	+9.61%	●	97.83%	+9.44%	●
Color Descriptors								
GCH [51]	Soccer	32.24%	34.71%	+7.66%	●	34.56%	+7.20%	●
ACC [52]	Soccer	37.23%	46.74%	+25.54%	●	47.52%	+27.64%	●
BIC [53]	Soccer	39.26%	47.99%	+22.24%	●	48.02%	+22.31%	●
Texture Descriptors								
LBP [54]	Brodatz	48.40%	50.26%	+3.84%	●	49.96%	+3.22%	●
CCOM [55]	Brodatz	57.57%	65.18%	+13.22%	●	65.26%	+13.36%	●
LAS [56]	Brodatz	75.15%	80.50%	+7.12%	●	79.60%	+5.92%	●

3.3.2. Color Image Retrieval

The experiments considering color image retrieval were conducted on a dataset [40] composed of images from 7 soccer teams, containing 40 images per class. Used descriptors include: Border/Interior Pixel Classification

(BIC) [53], Auto Color Correlograms (ACC) [52], and Global Color Histogram (GCH) [51].

Results considering MAP as score are presented in Table 3, for both distance and rank correlation measures. Significant positive gains can also be observed for all color descriptors, ranging from +7.66% to +25.54% for the Pearson and from +7.20% to +27.64% for the RBO measure.

3.3.3. Texture Retrieval

The popular Brodatz [41] dataset was used for experiment on texture retrieval. The dataset is composed of 1,776 images, in which 111 different textures are divided into 16 blocks. Three texture descriptors were considered: Local Binary Patterns (LBP) [54], Color Co-Occurrence Matrix (CCOM) [55], and Local Activity Spectrum (LAS) [56].

Results considering MAP scores are presented in Table 3. Considering the Pearson Correlation Coefficient, the gains ranged from +3.84% to +13.22%, while considering the RBO measure, the gains ranged from +3.22% to +13.36%. Notice that the LBP [54] descriptor, which presented the smaller gain among all descriptor (+3.84% and +3.22%), constitutes a challenging scenario, in which recent approaches [24, 38] presented negative gains.

3.3.4. Distance Fusion

We also evaluate the use of *Correlation Graph* for distance fusion, aiming at combining different CBIR descriptors. Two descriptors were selected for each visual property, including shape, color, and texture features. We selected the descriptors with similar and highest retrieval scores in distance learning tasks.

Table 4 presents results of MAP score of these descriptors. We can observe that significant gains are obtained when compared with the results of descriptors in isolation. For shape descriptors, the fused MAP score achieves **99.52%**, while the best descriptor considered yields 85.71%. Similar positive results are also presented for color and texture descriptors. For color descriptors, while the best descriptor in isolation yields only 39.26%, the distance fusion achieves a MAP of **49.12%**.

Table 4: Correlation Graph for distance fusion on shape, color and texture retrieval.

Descriptor	Type	Dataset	Correlation Measure	Score (MAP)
CFD [50]	Shape	MPEG-7	-	80.71%
ASC [45]	Shape	MPEG-7	-	85.28%
CFD+ASC	Shape	MPEG-7	Pearson	**99.52%**
CFD+ASC	Shape	MPEG-7	RBO	99.46%
ACC [52]	Color	Soccer	-	37.23%
BIC [53]	Color	Soccer	-	39.26%
BIC+ACC	Color	Soccer	Pearson	46.88%
BIC+ACC	Color	Soccer	RBO	**49.12%**
CCOM [55]	Texture	Brodatz	-	57.57%
LAS [56]	Texture	Brodatz	-	75.15%
LAS+CCOM	Texture	Brodatz	Pearson	83.66%
LAS+CCOM	Texture	Brodatz	RBO	**84.56%**

3.4. Object Retrieval

We also evaluated the proposed manifold learning algorithm for object retrieval tasks. The ETH-80 [42] dataset was considered for the experiment. The dataset is composed of 3,280 images of 128×128 pixels, and each image contains one single object. This dataset is equally divided into 8 classes where each class represents a different object.

The experiments were conducted considering four color descriptors: Border/Interior Pixel Classification (BIC) [53], Auto Color Correlograms (ACC) [52], Global Color Histogram (GCH) [51], and Color Structure Descriptor (CSD) [57].

Table 5 presents the MAP scores of each descriptor. Positive gains with statistical significance were also obtained for all descriptors. The gains range from +7.67% to +18.43%, considering the Pearson Correlation Coefficient and from +4.08% to +19.73% considering the RBO measure.

Table 5: Correlation Graph Distance for Object Retrieval on ETH-80 [42] dataset.

Descriptor	Original Score (MAP)	Correlation Graph Dist. Pearson	Gain	Statistic Signific. 99%	Correlation Graph Dist. RBO	Gain	Statistic Signific. 99%
BIC [53]	49.72%	55.79%	+12.21%	●	57.05%	+14.74%	●
ACC [52]	48.50%	52.22%	+7.67%	●	50.48%	+4.08%	●
CSD [57]	48.46%	57.39%	+18.43%	●	53.08%	+9.53%	●
GCH [51]	41.62%	47.26%	+13.55%	●	49.83%	+19.73%	●

3.5. Natural Image Retrieval

The University of Kentucky Recognition Benchmark - UKBench [44] has a total of 10,200 images. The dataset is composed of 2,550 objects or scenes, where each object/scene is captured 4 times from different viewpoints, distances and illumination conditions. Thus, it consists of 2,550 image classes, and each class has only 4 images. For the UKBench dataset, the same parameters of other datasets are used, except for $k = 5$ (for RBO) and $t_{start} = 0.5$ (for Pearson) are used, due to the very small number of similar images.

The experiments consider seven descriptors, exploiting various different features.[1] In the following, we briefly describe the descriptors used in the experiment:

- **Global Color:** Auto Color Correlogram (ACC) [52], Auto Color Correlogram Spatial Pyramid (ACC-SPy) [58], Scalable Color Descriptor (SCD) [59].

- **Global Color and Texture:**, Color and Edge Directivity Descriptor Spatial Pyramid [60, 58] (CEED-SPy), Fuzzy Color and Texture Histogram Spatial Pyramid [61, 58] (FCTH-SPy), Joint Composite Descriptor Spatial Pyramid [62, 58] (JCD-SPy).

- **Bag of Visual Words:** we considered a variant of vocabulary tree based retrieval (VOC) [44, 63], which uses SIFT features. For VOC, we considered the rank positions provided by recent approaches [37, 34][2] as the distances among images.[3]

- **Convolutional Neural Network (CNN):** features are extracted from the 7th layer using the Caffe framework [64]. A 4096-dimensional CNN-Caffe descriptor was considered for each input image resized of 256×256 pixels.

---

[1]The global descriptors were implemented based on the LIRE library – `http://www.semanticmetadata.net/lire/` (As of September 2015).

[2]`http://research.rutgers.edu/~shaoting/image_search.html` (As of September 2015).

[3]Images not present in the provided rankings had their distance defined as a constant $n_s = 200$ and only the RBO ranking function is considered.

Table 6 present the results of Correlation Graph for distance learning and distance fusion tasks. For evaluation purposes, the N-S score is used computed between 1 and 4, which corresponds to the number of relevant images among the first four image returned. Therefore, the highest achievable score is 4, indicating that all similar images are retrieved at top positions.

Very expressive gains can be observed. The N-S score obtained for the CNN-based descriptor, for example, was improved from 3.31 to 3.57, considering the RBO measure. For distance fusion tasks, the results are even more expressive reaching a very high score of **3.86**.

Table 6: Correlation Graph on the UKBench dataset (N-S score).

Descriptor	Original N-S Score	Correlation Graph Dist. Pearson	Gain	Correlation Graph Dist. RBO	Gain
CEED-SPy [58]	2.81	2.96	+5.34%	3.02	+7.47%
FCTH-SPy [58]	2.91	3.02	+3.78%	3.12	+7.22%
SCD [59]	3.15	3.18	+0.95%	3.33	+5.71%
ACC-SPy [58]	3.25	3.42	+5.23%	3.46	+6.46%
CNN-Caffe[64]	3.31	3.56	+7.55%	3.57	+7.85%
ACC [58]	3.36	3.49	+3.87%	3.56	+5.95%
VOC [63]	**3.54**	-	-	**3.73**	+5.37%
VOC [63]+ACC [52]	-	-	-	3.84	+8.47%
VOC [63]+CNN[64]	-	-	-	3.82	+7.91%
ACC [52]+CNN[64]	-	3.71	+4.80%	3.78	+6.78%
VOC [63]+ACC [52]+CNN[64]	-	-	-	**3.86**	+9.04%

3.6. Multimodal Image Retrieval

The UW dataset [43] was created at the University of Washington and consists of a roughly categorized collection of 1,109 images. This dataset includes vacation pictures from various locations. The images are partly annotated using keywords. The number of words per image ranges from 1 to 22, containing 6 words on the average. The dataset is classified into 18 categories, ranging from 22 images to 255 images per category.

The experiments consider twelve descriptors, which are listed below:

- **Visual Color Descriptors:** we considered three color descriptors on experiments: Border/Interior Pixel Classification (BIC) [53], Global Color Histogram (GCH) [51], and the Joint Autocorrelogram (JAC) [65].

- **Visual Texture Descriptors:** for texture we used the Homogeneous Texture Descriptor (HTD) [66], Quantized Compound Change Histogram (QCCH) [67], and Local Activity Spectrum (LAS) [56].

- **Textual Descriptors:** six well-known text similarity measures are considered for textual retrieval, like the Cosine similarity measure (COS), Term Frequency - Inverse Document Frequency (TF-IDF), and the Dice coefficient (DICE).

Table 7 presents the MAP scores for descriptors in isolation. Table 8 presents the results for the Correlation Graph method in multimedia retrieval tasks. The experiments were conducted considering four different scenarios: using all descriptors of each modality and using only the best descriptors.

Two baselines are also considered in the experiments: the traditional Borda [68] method and the recently proposed Reciprocal Rank Fusion [69]. It can be observed that, except for the combination of all visual descriptors, all the remaining results overcome the best individual descriptor (52.26%). The best multimodal retrieval result (**75.59%**) presents a very significant gain of **+44.64%** over the best individual descriptor in isolation.

3.7. Comparison with Other Approaches

The proposed Correlation Graph method were also evaluated in comparison with other state-of-the-art post-processing and retrieval methods. Two well-known datasets commonly used for benchmark were considered. The comparisons are presented in next sub-sections.

3.7.1. Shape Retrieval

The MPEG-7 dataset [39] was considered for the first comparison, since it has been commonly used in the evaluation and comparison of post-processing and distance fusion approaches. The bull's eye score (recall@40) is used as evaluation measure.

Table 9 presents the bull's eye score obtained by the Correlation Graph Distance in comparison with several other post-processing methods recently proposed.

For distance learning, the three best results of the proposed approach are reported. For distance fusion, the CFD [50]+ASC [45] combination is

Table 7: Correlation Graph for visual and textual retrieval on the UW dataset [43].

Descriptor	Type	Original Score (MAP)
GCH [51]	Visual/Color	31.75%
BIC [53]	Visual/Color	43.46%
JAC [65]	Visual/Color	**52.26%**
QCCH [67]	Visual/Texture	17.81%
LAS [56]	Visual/Texture	20.44%
HTD [66]	Visual/Texture	22.61%
DICE [70]	Textual	50.73%
OKAPI [71]	Textual	**51.68%**
BOW [72]	Textual	48.84%
COS [73]	Textual	41.80%
JACKARD [70]	Textual	50.29%
TF-IDF [73]	Textual	49.25%

Table 8: Correlation Graph on multimodal retrieval tasks (MAP as score).

Retrieval Task	Descriptors	Correlation Graph Dist. Pearson	Correlation Graph Dist. RBO	Baselines Borda [68]	Baselines Reciprocal Fusion [69]
Visual	All visual descriptors	47.22%	49.77%	40.29%	43.29%
Textual	All textual descriptors	60.46%	61.45%	53.07%	53.14%
Multimodal	All descriptors	**69.87%**	**73.25%**	54.89%	59.34%
Visual	BIC [53]+JAC [65]	60.65%	60.38%	52.54%	53.00%
Textual	DICE [70]+OKAPI [71]	63.33%	62.28%	54.57%	54.31%
Multimodal	BIC [53]+JAC [65]+ DICE [70]+OKAPI [71]	**74.79%**	**75.59%**	61.91%	63.67%

Table 9: Comparison with various 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 [74]	-	80.03%
CFD [50]	-	84.43%
IDSC [49]	-	85.40%
SC [75]	-	86.80%
ASC [45]	-	88.39%
AIR [46]	-	93.67%
Unsupervised Post-Processing Methods		
Graph Transduction [27]	IDSC [49]	91.00%
LCDP [15]	IDSC [49]	93.32%
Shortest Path Propagation [26]	IDSC [49]	93.35%
Mutual kNN Graph [16]	IDSC [49]	93.40%
Pairwise Recommendation [25]	ASC [45]	94.66%
RL-Sim [24]	ASC [45]	94.69%
Reciprocal kNN Manifold [29]	ASC [45]	95.27%
Cor. Graph Distance - Pearson	**ASC [45]**	**95.50%**
LCDP [15]	ASC [45]	95.96%
Tensor Product Graph [17]	ASC [45]	96.47%
RL-Sim [24]	AIR [46]	99.94%
Reciprocal kNN Manifold [29]	AIR [46]	99.94%
Cor. Graph Distance - RBO	**AIR [46]**	**99.95%**
Tensor Product Graph [17]	AIR [46]	99.99%
Generic Diffusion Process [28]	AIR [46]	100%
Cor. Graph Distance - Pearson	**AIR [46]**	**100%**
Unsupervised Distance Fusion Methods		
Reciprocal Rank Fusion [69]	CFD+IDSC	94.98%
Reciprocal Rank Fusion [69]	CFD+ASC	96.25%
Co-Transduction [76]	IDSC [49]+DDGM [74]	97.31%
Self-Smoothing Operator [12]	SC [75] +IDSC [49]	97.64%
Self-Smoothing Operator [12]	SC [75] +IDSC [49] +DDGM [74]	99.20%
Pairwise Recommendation [25]	CFD [50]+IDSC [49]	99.52%
RL-Sim [24]	CFD [50]+ASC [45]	99.65%
Cor. Graph Distance - RBO	**CFD [50]+ASC [45]**	**99.80%**
Cor. Graph Distance - Pearson	**CFD [50]+ASC [45]**	**99.92%**

considered. The *Correlation Graph Distance* presents comparable and better effectiveness performance, achieving a bull's eye score of **100%** for the AIR [46] shape descriptor.

3.7.2. Natural Image Retrieval

The UKBench is a well-known image dataset commonly used as benchmark for image retrieval methods. The UKBench is a very challenging dataset specially for unsupervised learning approaches, due to the small number of images in each class.

Table 10 presents the N-S scores obtained by various recent retrieval state-of-the-art methods, including fusion approaches. The best scores obtained by the Correlation Graph Distance are also reported. As we can observe, the *Correlation Graph Distance* presents comparable and better effectiveness in comparison with various approaches. The proposed approach reached a very high N-S score of **3.86** for RBO correlation measure fusing VOC+ACC+CNN.

Table 10: Retrieval comparison between the proposed Correlation Graph algorithm and recent retrieval methods on the the UKBench dataset.

N-S scores for recent retrieval methods						
Luo *et al.* [77]	Zheng *et al.* [78]	Qin *et al.* [33]	Jégou *et al.* [79]	Wang *et al.* [80]	Zhang *et al.* [37]	Zheng *et al.* [36]
3.56	3.57	3.67	3.68	3.68	3.83	3.84

N-S scores for the proposed Correlation Graph method				
C. Graph RBO VOC	C. Graph RBO ACC+CNN	C. Graph RBO VOC+CNN	C. Graph RBO VOC+ACC	**C. Graph RBO VOC+ACC+CNN**
3.73	3.78	3.82	3.84	**3.86**

4. Conclusions

An unsupervised manifold learning approach for improving image retrieval tasks is discussed in this paper. The algorithm performs a correlation analysis for constructing a graph representation of the dataset. The Correlation Graph (CG) and Strongly Connected Components (SCCs) are used for discovering the intrinsic geometry of the dataset manifold, improving distance among images.

The algorithm is able to exploit both distance and rank correlation measures for constructing the Correlation Graph. The distance correlation is measured by the Pearson correlation coefficient, while the rank correlation analysis is performed using a recent proposed rank correlation measure, the Rank-Biased Overlap [31] (RBO). The use of rank information enables the construction of the graph representation without the need of distance scores.

A large set of experiments was conducted for assessing the effectiveness of the proposed approach, considering different descriptors and datasets. The high effectiveness of the manifold learning algorithm is demonstrated by the

30

experimental results obtained in several image retrieval tasks. The effectiveness gains associated with the low computational efforts required represent a significant advantage of the discussed method when compared with existing approaches proposed in the literature.

Future work focuses on: (i) the investigation of relevance feedback and collaborative image retrieval using the correlation graph; and (ii) the evaluation of efficiency and scalability aspects using parallel and heterogeneous computing scenarios.

Acknowledgments

References

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

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

[3] Y. Yan, E. Ricci, R. Subramanian, G. Liu, N. Sebe, Multitask linear discriminant analysis for view invariant action recognition, IEEE Transactions on Image Processing 23 (12) (2014) 5599–5611.

[4] Y. Yan, E. Ricci, R. Subramanian, G. Liu, O. Lanz, N. Sebe, A multi-task learning framework for head pose estimation under target motion, IEEE Transactions on Pattern Analysis and Machine Intelligence Online, To appear.

[5] R. Datta, D. Joshi, J. Li, J. Z. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys 40 (2) (2008) 5:1–5:60.

[6] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.

[7] Y. Yan, Y. Yang, D. Meng, G. Liu, W. Tong, A. Hauptmann, N. Sebe, Event oriented dictionary learning for complex event detection, IEEE Transactions on Image Processing 24 (6) (2015) 1867–1878.

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

[9] S. Lafon, A. B. Lee, Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (9) (2006) 1393–1403.

[10] J. B. Tenenbaum, V. d. Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[11] S. Tan, L. Liu, C. Peng, L. Shao, Image-to-class distance ratio: A feature filtering metric for image classification, Neurocomputing On-Line, To appear.

[12] 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.

[13] Z. Cui, H. Chang, S. Shan, X. Chen, Generalized unsupervised manifold alignment, in: Advances in Neural Information Processing (NIPS'2014), 2014, pp. 2429–2437.

[14] D. Zhou, J. Weston, A. Gretton, O. Bousquet, B. Schlkopf, Ranking on data manifolds, in: Advances in Neural Information Processing Systems (NIPS'2004), 2004, pp. 169 – 176.

[15] X. Yang, S. Koknar-Tezel, L. J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval., in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2009), 2009, pp. 357–364.

[16] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: Asian Conference on Computer Vision (ACCV'2009), 2009, pp. 655–666.

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

[18] E. de Ves, G. Ayala, X. Benavent, J. Domingo, E. Dura, Modeling user preferences in content-based image retrieval: A novel attempt to bridge the semantic gap, Neurocomputing 168 (2015) 829 – 845.

[19] Y.-Y. Lin, T.-L. Liu, H.-T. Chen, Semantic manifold learning for image retrieval, in: Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA '05, 2005, pp. 249–258.

[20] Q. Tian, J. Yu, Q. Xue, N. Sebe, A new analysis of the value of unlabeled data in semi-supervised learning for image retrieval, in: Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, Vol. 2, 2004, pp. 1019–1022 Vol.2.

[21] H. Berlin, The neural basis of the dynamic unconscious, Neuropsychoanalysis 13 (1) (2011) 63–71.

[22] R. M. Warren, Perceptual restoration of missing speech sounds, Science 167 (1970) 392–393.

[23] L. Mlodnow, Subliminal: How Your Unconscious Mind Rules Your Behavior, Vintage, 2013.

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

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

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

[27] X. Yang, X. Bai, L. J. Latecki, Z. Tu, Improving shape retrieval by learning graph transduction, in: European Conference on Computer Vision (ECCV'2008), Vol. 4, 2008, pp. 788–801.

[28] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, in: IEEE Conferece on Computer Vision and Pattern Recognition (CVPR'2013), 2013, pp. 1320–1327.

[29] 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 and Vision Computing 32 (2) (2014) 120 – 130.

[30] 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.

[31] W. Webber, A. Moffat, J. Zobel, A similarity measure for indefinite rankings, ACM Transactions on Information Systems 28 (4) (2010) 20:1–20:38.

[32] 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.

[33] D. Qin, S. Gammeter, L. Bossard, T. Quack, L. van Gool, Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011), 2011, pp. 777 –784.

[34] S. Zhang, M. Yang, T. Cour, K. Yu, D. N. Metaxas, Query specific fusion for image retrieval, in: ECCV, 2012, pp. 660–673.

[35] R. Tarjan, Depth first search and linear graph algorithms, SIAM Journal on Computing.

[36] 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.

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

[38] D. C. G. Pedronette, O. A. B. Penatti, R. T. Calumby, R. da S. Torres, Unsupervised distance learning by reciprocal knn distance for image retrieval, in: International Conference on Multimedia Retrieval (ICMR'14), 2014.

[39] L. J. Latecki, R. Lakmper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2000), 2000, pp. 424–429.

[40] J. van de Weijer, C. Schmid, Coloring local feature extraction, in: European Conference on Computer Vision (ECCV'2006), Vol. Part II, 2006, pp. 334–348.

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

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

[43] T. Deselaers, D. Keysers, H. Ney, Features for image retrieval: an experimental comparison, Information Retrieval 11 (2) (2008) 77–107.

[44] D. Nistér, H. Stewénius, Scalable recognition with a vocabulary tree, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2006), Vol. 2, 2006, pp. 2161–2168.

[45] H. Ling, X. Yang, L. J. Latecki, Balancing deformability and discriminability for shape matching, in: European Conference on Computer Vision (ECCV'2010), Vol. 3, 2010, pp. 411–424.

[46] R. Gopalan, P. Turaga, R. Chellappa, Articulation-invariant representation of non-planar shapes, in: 11th European Conference on Computer Vision (ECCV'2010), Vol. 3, 2010, pp. 286–299.

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

[48] N. Arica, F. T. Y. Vural, BAS: a perceptual shape descriptor based on the beam angle statistics, Pattern Recognition Letters 24 (9-10) (2003) 1627–1639.

[49] H. Ling, D. W. Jacobs, Shape classification using the inner-distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2) (2007) 286–299.

[50] D. C. G. Pedronette, R. da S. Torres, Shape retrieval using contour features and distance optmization, in: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP'2010), Vol. 1, 2010, pp. 197 – 202.

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

[52] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlograms, in: IEEE Conference on Computer Vision and Pattern Recognition (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: ACM Conference on Information and Knowledge Management (CIKM'2002), 2002, pp. 102–109.

[54] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.

[55] V. Kovalev, S. Volmer, Color co-occurence descriptors for querying-by-example, in: International Conference on Multimedia Modeling, 1998, p. 32.

[56] B. Tao, B. W. Dickinson, Texture recognition and image retrieval using gradient indexing, Journal of Visual Comunication and Image Representation 11 (3) (2000) 327–342.

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

[58] M. Lux, Content based image retrieval with LIRe, in: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, 2011.

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

[60] 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, ICVS'08, 2008, pp. 312–322.

[61] 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.

[62] 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.

[63] 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.

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

[65] A. Williams, P. Yoon, Content-based image retrieval using joint correlograms, Multimedia Tools and Applications 34 (2) (2007) 239–248.

[66] P. Wu, B. S. Manjunanth, S. D. Newsam, H. D. Shin, A texture descriptor for image retrieval and browsing, in: IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'99), 1999, pp. 3–7.

[67] C.-B. Huang, Q. Liu, An orientation independent texture descriptor for image retrieval, in: International Conference on Communications, Circuits and Systems (ICCCAS 2007), 2007, pp. 772–776.

[68] H. P. Young, An axiomatization of borda's rule, Journal of Economic Theory 9 (1) (1974) 43–52.

[69] G. V. Cormack, C. L. A. Clarke, S. Buettcher, Reciprocal rank fusion outperforms condorcet and individual rank learning methods, in: ACM SIGIR Conference on Research and Development in Information Retrieval, 2009, pp. 758–759.

[70] J. Lewis, S. Ossowski, J. Hicks, M. Errami, H. R. Garner, Text similarity: an alternative way to search medline, Bioinformatics 22 (18) (2006) 2298–2304.

[71] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, M. Gatford, Okapi at trec-3, in: Text REtrieval Conference, 1994, pp. 109–126.

[72] M. Carrillo, E. Villatoro-Tello, A. López-López, C. Eliasmith, M. Montes-Y-Gómez, L. Villaseñor Pineda, Representing context information for document retrieval, in: 8th International Conference on Flexible Query Answering Systems (FQAS'09), 2009, pp. 239–250.

[73] R. A. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[74] Z. Tu, A. L. Yuille, Shape matching and recognition - using generative models and informative features, in: European Conference on Computer Vision (ECCV'2004), 2004, pp. 195–209.

[75] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.

[76] X. Bai, B. Wang, X. Wang, W. Liu, Z. Tu, Co-transduction for shape retrieval, in: European Conference on Computer Vision (ECCV'2010), Vol. 3, 2010, pp. 328–341.

[77] Q. Luo, S. Zhang, T. Huang, W. Gao, Q. Tian, Superimage: Packing semantic-relevant images for indexing and retrieval, in: International Conference on Multimedia Retrieval (ICMR '14), 2014, pp. 41:41–41:48.

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

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

[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.