This paper presents a novel manifold learning approach that takes into account the intrinsic dataset geometry. The dataset structure is modeled in terms of a Correlation Graph and analyzed using Strongly Connected Components (SCCs). The proposed manifold learning approach defines a more effective distance among images, used to improve the effectiveness of image retrieval systems. Several experiments were conducted for different image retrieval tasks involving shape, color, and texture descriptors. The proposed approach yields better results in terms of effectiveness than various methods recently proposed in the literature.

Index Terms—content-based image retrieval, unsupervised manifold learning, correlation graph

2. MANIFOLD LEARNING BY CORRELATION GRAPH

Our objective is to represent the intrinsic geometry of a dataset manifold in terms of a distance correlation analysis. In this way, we propose a graph-based approach that can be roughly divided into three main steps: first, the distance correlation between each dataset image and the images placed at top positions of its ranked list is computed. For each ranked list, only a small set of images (which are the most likely to be similar to the query image) are selected based on a correlation threshold. Selected images have edges added to a Correlation Graph. The Correlation Graph is then analyzed for identify-
ing Strongly Connected Components (SCCs). SCCs are used for expanding the set of similar images by taking account the intrinsic geometry of the dataset manifold. Finally, the first two steps are repeated using different values of correlation threshold. For each value, the edges of the Correlation Graph and the identified SCCs are used to compute a novel distance called Correlation Graph Distance.

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 (marked with a triangle) in each moon. In the following, all other data points are assigned to the closest labeled point, determining their color. 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 gray. 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.

2.1. Correlation Graph

Let \( C = \{img_1, img_2, \ldots, img_n\} \) be an image collection, where \( n \) is the size of the collection. Let \( \rho(i, j) \) denotes the distance between two images \( img_i \) and \( img_j \), according to a given image descriptor. Let \( \tau_q = \{img_1, img_2, \ldots, img_{n_s}\} \) be a ranked list, which can be defined as a permutation of the subset \( C_s \subset C \). The subset \( C_s \) contains the \( n_s \) most similar images to query image \( img_q \), such that \( |C_s| = n_s \). We interpret \( \tau_q(i) \) as the position (or rank) of image \( img_i \) in the ranked list \( \tau_q \), computed in response to the query image \( img_q \).

Given a directed graph \( G = (V, E) \), the set of vertices \( V \) is defined by the image collection \( C \), such that each image is represented by a node and \( V = C \). The edge set \( E \) is defined considering the distances correlation among images at the top \( n_s \) positions of each ranked list, as follows: \( E = \{(img_q, img_j) \mid \tau_q(j) \leq n_s \land \text{cor}(q, j) \geq t_c\} \), where \( \text{cor}(q, j) \) is the correlation score between \( img_q \) and \( img_j \) and \( t_c \) is the correlation threshold considered. Therefore, there will be edge from \( img_q \) to \( img_j \), if: (i) \( img_j \) is at the top positions of ranked of \( img_q \); and (ii) the distance correlation between them are greater than a given threshold \( t_c \).

The correlation score \( \text{cor}(q, j) \) is computed by the Pearson’s Correlation Coefficient, considering the distances to the \( k \)-nearest neighbors of \( img_q \) and \( img_j \). Let \( N_k(q) \) be the set containing the \( k \)-nearest neighbors to given image \( img_q \). Let \( N_k(q, j) \) be the union set containing the \( k \)-nearest neighbors of both images \( img_q \) and \( img_j \), such that \( N_k(q, j) = N_k(q) \cup N_k(j) \). We define two vectors \( X \) and \( Y \) containing, respectively, the distances from images \( img_q \) and \( img_j \) to each image \( img_k \in N_k(q, j) \). Let \( img_k \) be the \( i \)-th image of the set \( N_k(q, j) \), we define \( X_i = \rho(q, i) \) and \( Y_i = \rho(j, i) \).

The correlation score \( \text{cor}(q, j) \) is defined as follows:

\[
\text{cor}(q, j) = \frac{\sum_{i=1}^{n_s} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n_s} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n_s} (Y_i - \bar{Y})^2}}
\] (1)

2.2. Strongly Connected Components

The edges defined by the Correlation Graph give a very strong indication of similarity among images. However, although very precise, the edges include a very small neighborhood (as can be observed in Figure 2). In this way, we aim at expanding the similarity neighborhood, but still considering the geometry of the dataset manifold. Recently, the reciprocal neighborhood [11, 12] has been considered for analyzing the dataset structure. With the same objective, we consider the Strongly Connected Components (SCCs) of the Correlation Graph. 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. Since the SCCs define reciprocal references among a set of nodes, it can be considered as an extension of the concept of reciprocal neighborhood. We used the Tarjan [13] algorithm for computing the SCCs, which is linear on the size of the graph. Each SCC is defined as a set of images \( S \). Therefore, the overall output of the algorithm is a set of SCCs \( S = \{S_1, S_2, \ldots, S_m\} \), which is used for computing the Correlation Graph Distance.

2.3. 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 more effective distance among images. In this way, we define a Correlation Graph Similarity Score $W_{i,j}$, which aims at quantifying the association between two given images $img_i, img_j$ according to the Correlation Graph and SCCs. The similarity score $W_{i,j}$ is defined in terms of increments, according to the Correlation Graph edges and SCCs. 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 two images that belong to a same SCC. Algorithm 1 outlines the proposed method for computing the similarity score $W_{i,j}$.

Algorithm 1 Correlation Graph Distance

Require: Correlation Graph $G = (V, E)$, Set of SCCs $S$
Ensure: Correlation Graph Similarity Score $W_{i,j}$
1. $t_c \leftarrow t_{start}$
2. while $t_c \leq 1$
3. { (Correlation Graph )
4. for all $img_q \in V$
5. for all $img_i, img_j \in E(q)$ do
6. $W_{i,j} \leftarrow W_{i,j} + t_c$
7. end for
8. end for
9. { Strongly Connected Components }
10. for all $S_c \in S$
11. for all $img_i, img_j \in S_c$ do
12. $W_{i,j} \leftarrow W_{i,j} + t_c$
13. end for
14. end for
15. $t_c \leftarrow t_c + t_{inc}$
16. end while

Finally, based on the similarity score $W_{i,j}$, we compute the Correlation Graph Distance $\rho_c(i,j)$ as follows:

$$\rho_c(i,j) = \frac{1}{1 + W_{i,j}}.$$  \hfill (2)

3. EXPERIMENTAL EVALUATION

This section presents a set of conducted experiments for assessing the effectiveness of the proposed method.

3.1. Impact of Parameters

This section aims at defining the best parameter setting. We conducted four experiments considering the MPEG-7 [14] dataset. The MPEG-7 [14] dataset is a well-known shape dataset, composed of 1400 shapes divided into 70 classes. The Mean Average Precision (MAP) was considered as effectiveness measure. First, we evaluate the impact of the parameters $k$ (size of the neighborhood set) and $t_{start}$ (start value of Correlation Threshold $t_c$). Figure 4 illustrates the values of MAP according to variations of $k$ and $t_{start}$ for the Aspect Shape Context (ASC) [15] descriptor. We can observe a large red region indicating high retrieval scores, which demonstrates the robustness of the proposed method. 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) [15] and Articulation-Invariant Representation (AIR) [16]. We observed that only a small subset of ranked lists ($n_s = 200$) is enough to achieve the best results. Different values of $t_{inc}$, in turn, did not impact the effectiveness of the descriptors. We used the values of $k = 25$, $t_{start} = 0.35$, $n_s = 200$, and $t_{inc} = 0.005$ for all datasets and different descriptors.

3.2. General Image Retrieval Tasks

This section presents the results of proposed method for general image retrieval tasks. We also conducted statistical paired $t$-tests, comparing the results before and after the use of the proposed manifold learning algorithm.

We evaluate the use of our method for shape retrieval using the MPEG-7 [14] dataset. Six shape descriptors were considered: Segment Saliences (SS) [17], Beam Angle Statistics (BAS) [18], Inner Distance Shape Context (IDSC) [19], Contour Features Descriptors (CFD) [20], Aspect Shape Context (ASC) [15], and Articulation-Invariant Representation (AIR) [16]. We considered two effectiveness measures for the MPEG-7 dataset: the Mean Average Precision (MAP) and the so-called Bull’s Eye Score, commonly used for this dataset. This score counts all matching objects within the 40 most similar candidates. Since each class consists of 20 objects, the retrieved score is normalized with the highest possible number of hits. Table 1 presents the MAP scores, while Table 2 presents the results considering the bull’s eye score of evaluated descriptors. Significant positive gains are observed for all descriptors, considering both measures, ranging from $+6.90\%$ to $+34.54\%$.

The experiments with color descriptor were conducted on a dataset [29] composed of images from 7 soccer teams, containing 40 images per class. Used descriptors include: Border/Interior Pixel Classification (BIC) [23], Auto Color Correlograms (ACC) [22], and Global Color Histogram (GCH) [21]. Table 1 presents the experimental results considering MAP as score. We can observe a positive gain for all color descriptors ranging from $+7.29\%$ to $+20.65\%$.

The experiments considering texture descriptors were conducted on the Brodatz [30] dataset, which is composed of 111 different textures. Each texture is divided into 16
Table 1. Correlation Graph Distance for various image retrieval tasks. Mean Average Precision considering shape, color, and texture descriptors.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Dataset</th>
<th>Score (MAP)</th>
<th>Correlation Graph Distance</th>
<th>Gain</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape Descriptors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS [17]</td>
<td>Brodatz</td>
<td>37.67%</td>
<td>50.68%</td>
<td>+34.54%</td>
<td>●</td>
</tr>
<tr>
<td>BAS [18]</td>
<td>Brodatz</td>
<td>71.52%</td>
<td>81.93%</td>
<td>+14.61%</td>
<td>●</td>
</tr>
<tr>
<td>IDSC [19]</td>
<td>Brodatz</td>
<td>81.70%</td>
<td>89.39%</td>
<td>+6.91%</td>
<td>●</td>
</tr>
<tr>
<td>CFD [20]</td>
<td>Brodatz</td>
<td>80.71%</td>
<td>91.93%</td>
<td>+13.90%</td>
<td>●</td>
</tr>
<tr>
<td>ASC [15]</td>
<td>Brodatz</td>
<td>85.28%</td>
<td>92.53%</td>
<td>+7.25%</td>
<td>●</td>
</tr>
<tr>
<td>AIR [16]</td>
<td>Brodatz</td>
<td>89.39%</td>
<td>97.98%</td>
<td>+9.61%</td>
<td>●</td>
</tr>
<tr>
<td><strong>Color Descriptors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCH [21]</td>
<td>Soccer</td>
<td>32.24%</td>
<td>34.59%</td>
<td>+7.29%</td>
<td>●</td>
</tr>
<tr>
<td>ACC [22]</td>
<td>Soccer</td>
<td>37.23%</td>
<td>45.24%</td>
<td>+21.51%</td>
<td>●</td>
</tr>
<tr>
<td>BIC [23]</td>
<td>Soccer</td>
<td>39.26%</td>
<td>47.37%</td>
<td>+20.65%</td>
<td>●</td>
</tr>
<tr>
<td><strong>Texture Descriptors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LBP [24]</td>
<td>Brodatz</td>
<td>48.40%</td>
<td>50.12%</td>
<td>+3.55%</td>
<td>●</td>
</tr>
<tr>
<td>CCOM [25]</td>
<td>Brodatz</td>
<td>57.57%</td>
<td>64.73%</td>
<td>+12.44%</td>
<td>●</td>
</tr>
<tr>
<td>LAS [26]</td>
<td>Brodatz</td>
<td>75.15%</td>
<td>79.87%</td>
<td>+4.68%</td>
<td>●</td>
</tr>
</tbody>
</table>

We also evaluated the proposed approach for object retrieval tasks. The experiments were conducted on the ETH-80 [27] dataset, which is composed of 3,280 images. Each image contains one single object. This dataset is equally divided into 8 classes where each class represents a different object, and all images have 128 × 128 pixels. Four color descriptors were used: BIC [23], ACC [22], GCH [21] and Color Structure Descriptor (CSD) [28]. Table 3 presents the MAP scores of each descriptor. Positive gains were obtained for all descriptors, ranging from +4.39% to +18.10%.

Finally, we also evaluate our method in comparison with other state-of-the-art post-processing methods. We consider again the MPEG-7 dataset [14], commonly used in the evaluation of post-processing methods. Table 4 presents the results of the proposed Correlation Graph Distance considering the Bull’s Eye Score in comparison with several other post-processing methods recently proposed. The Correlation Graph Distance presents comparable and better effectiveness performance when compared to various recently proposed methods. Note that the Correlation Graph Distance achieves a Bull’s Eye Score of 100% for the AIR [16] shape descriptor.

4. CONCLUSIONS

We have presented a novel unsupervised manifold learning method for improving image retrieval tasks. The proposed approach exploits the distance correlation for constructing a graph representation of the dataset. Based on the graph, strongly connected components are used for discovering the intrinsic geometry of the dataset manifold. We conducted a large set of experiments for assessing the effectiveness of the proposed approach, considering different descriptors and datasets. Experimental results demonstrated the high effectiveness of our method in several image retrieval tasks. Future work focuses on: (i) the investigation of distance fusion approaches for descriptor combination; and (ii) the evaluation of efficiency and scalability aspects.

5. 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.
6. REFERENCES


[29] Joost van de Weijer and Cordelia Schmid, “Coloring local feature extraction,” in ECCV.

