Unsupervised Distance Learning By Reciprocal kNN Distance for Image Retrieval

Daniel C. G. Pedronette[1], Otávio A. B. Penatti[2,3], Rodrigo T. Calumby[2,4], Ricardo da S. Torres[2]
[1]Department of Statistic, Applied Math. and Computing, Universidade Estadual Paulista (UNESP), Rio Claro, SP, 13506-900, Brazil
[2]RECOD Lab, Institute of Computing, University of Campinas (UNICAMP), Campinas, SP, 13083-852, Brazil
[3]Advanced Technologies, SAMSUNG Research Institute, Campinas, SP, 13097-104, Brazil
[4]Department of Exact Sciences, University of Feira de Santana (UEFS), Feira de Santana, BA, 44036-900, Brazil
daniel@rc.unesp.br, o.penatti@samsung.com, rtcalumby@ecomp.uefs.br, rtorres@ic.unicamp.br

ABSTRACT

This paper presents a novel unsupervised learning approach that takes into account the intrinsic dataset structure, which is represented in terms of the reciprocal neighborhood references found in different ranked lists. The proposed *Reciprocal kNN Distance* defines a more effective distance between two images, and is 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 is also evaluated on multimodal retrieval tasks, considering visual and textual descriptors. Experimental results demonstrate the effectiveness of proposed approach. The *Reciprocal kNN Distance* yields better results in terms of effectiveness than various state-of-the-art algorithms.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Search process

General Terms

Experimentation, Performance

Keywords

Content-based image retrieval, Unsupervised distance learning

1. INTRODUCTION

The goal of Content-Based Image Retrieval (CBIR) systems is to retrieve the most similar images in a collection by taking into account image visual properties [10]. The distance among images are computed according to a given image descriptor and the collection images are ranked in decreasing order of similarity. Therefore, the effectiveness of CBIR systems is very dependent on the distance measure adopted. For decades, several different visual features have been proposed for image retrieval tasks (based on shape, color, and texture properties).

More recently, however, aiming at improving the retrieval effectiveness of CBIR systems, the research community has also focused on other stages of the retrieval pipeline, which are not directly related to low-level feature extraction [26]. One of the techniques adopted consists in using unsupervised approaches to associate low-level features with query patterns. The use of unsupervised approaches presents the important advantage of not requiring any training or labeled data.

CBIR systems often compute only pairwise distance among images, and therefore, ignore the information encoded in the relationship among images in a given collection. This important source of information, commonly refereed as contextual information [16], can be exploited by various techniques aiming at improving the effectiveness of CBIR systems without the need of user intervention. The objective of such initiatives is to capture and utilize the intrinsic structure of the relationships among images in a collection [17].

In this paper, we propose a novel unsupervised learning approach called *Reciprocal kNN Distance*, which aims at exploiting the intrinsic collection structure by analyzing reciprocal neighborhoods. The objective of *Reciprocal kNN Distance* is to define a more effective distance between two images by analyzing the reciprocal references among images at top positions of their ranked lists. The main novelty of the proposed approach consists in the combination of two techniques that have been receiving great attention on post-processing methods recently: (*i*) the similarity between ranked lists [6, 32], and (*ii*) the reciprocal neighborhood [33, 51]. The *Reciprocal kNN Distance* models the similarity between ranked lists in terms of the density of reciprocal neighborhoods.

A large experimental evaluation was conducted, considering different datasets, image descriptors, and retrieval tasks. Experiments were conducted on three image datasets considering twelve different visual descriptors (shape, color, and texture descriptors). The proposed approach was also evaluated on object retrieval and multimodal image retrieval tasks, considering various visual and textual descriptors. The experimental evaluation demonstrates that the proposed method can achieve significant effectiveness improvements in several image retrieval tasks. We also evaluated the proposed *Reciprocal kNN Distance* in comparison with several other state-of-the-art approaches considering a shape dataset commonly used for benchmarking. Experimental results demonstrate that the proposed unsupervised learning approach yields better results in terms of effectiveness performance than various methods recently proposed in the literature.

The paper is organized as follows: Section 2 discusses related work and Section 3 presents the problem formulation. In Section 4, we present the *Reciprocal kNN Distance* learning method. Section 5 presents the experimental evaluation and, finally, Section 6 presents our conclusions and possible future work.

2. RELATED WORK

Although effective, supervised image retrieval techniques often require very expensive human efforts for obtaining training or labeled data. In fact, unlabeled data is far easier to obtain, and therefore unsupervised learning represents a very attractive solution for many practical situations. In a sense, "unsupervised learning can be thought of as finding patterns in the data above and beyond what would be considered unstructured noise" [12]. Given that it does not require any feedback or user intervention, the goal of these approaches is to build representations of the input space that can be used for predicting future inputs, distance learning, or even ranking dataset objects.

Two classic examples of unsupervised techniques are clustering and dimensionality reduction. When dealing with data in high-dimensional spaces, a challenging problem is how to reduce the complexity of a data set preserving information that is important for understanding the data structure itself. That is also valid for performing tasks such as clustering, classification, and regression [20]. The dimensionality reduction term designates methods that aim at finding meaningful low-dimensional structures hidden in their high-dimensional observations [39].

In the information retrieval applications, unsupervised learning approaches have also been proposed aiming at improving the effectiveness of distance measures. The term *"global ranking"* was introduced in [34]. Basically, a global ranking approach considers that it is better to define the ranking model as a function of all the objects to be ranked, by taking into account the relationships among objects.

Various approaches have also been proposed aiming at improving the distance measures in CBIR systems [18, 45, 48, 49]. The objective of these approaches consists in replacing pairwise similarities by more global affinity measures [49]. Usually, they exploit the information encoded in the relationships among images, commonly referred to as contextual information. These methods often use iterative strategies to process contextual information. Various different techniques have been employed, such as clustering [30], graphs [18], and diffusion process [48, 49].

More recently, distance learning approaches based on the similarity between ranked lists [6, 32] have been proposed. These methods are based on the conjecture that contextual information is encoded in the ranked lists. Other recent approaches are based on the concept of reciprocal neighborhood [33, 51] as a stronger indicator of similarity between images.

The proposed *Reciprocal kNN Distance* aims at combining the analysis of ranked lists and reciprocal neighborhoods by modelling the similarity between ranked lists in terms of the density of reciprocal references found in ranked lists. Another novelty of the proposed approach relies on the fact that the *Reciprocal kNN Distance* is non-iterative and, therefore, requires no parameter for the number of iterations or convergence criterion as iterative methods recently proposed [32, 47]. In this way, it also reduces the computational efforts needed for distance learning.

3. PROBLEM FORMULATION

Let $\mathcal{C}$=$\{img_1, img_2, \ldots, img_n\}$ be an image collection. Let $n = |\mathcal{C}|$ denotes the size of the collection $\mathcal{C}$. Let $\mathcal{D}$ be an image descriptor which can be defined [8] 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, i.e., the distance distance between two images img_i and img_j is given by the value of $\rho(\epsilon(img_i), \epsilon(img_j))$. For simplicity and readability purposes, we use the notation $\rho(i, j)$ along the paper.

The distance $\rho(i, j)$ among all images img_i, $img_j \in \mathcal{C}$ can be computed to obtain a squared $n \times n$ distance matrix A, such that $A_{ij} = \rho(i, j)$. Also 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=$(img_1, img_2, \ldots, img_n)$ 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 $[N] = \{1, 2, \ldots, n\}$. For a permutation τ_q, we interpret $\tau_q(i)$ as the position (or rank) of image img_i in the ranked list τ_q. We can say that, if img_i is ranked before img_j in the ranked list of img_q, that is, $\tau_q(i) < \tau_q(j)$, then $\rho(q, i) \leq \rho(q, j)$. We also can take every image $img_i \in \mathcal{C}$ as a query image img_q, in order to obtain a set $\mathcal{R} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ of ranked lists for each image of the collection $\mathcal{C}$.

Our problem consists in redefining the distance ρ by computing a more effective distance function ρ_r. The objective of function $\rho_r(i, j, \mathcal{R})$ is to exploit the information encoded in the set $\mathcal{R}$ by analyzing the reciprocal references found in the top positions of ranked lists aiming at improving the effectiveness of distances among images. In a math notation: ρ_r: $\mathcal{C} \times \mathcal{C} \times \mathcal{R} \rightarrow \mathbb{R}$ is a distance function between two images img_i, $img_j \in \mathcal{C}$ that considers the information of the set of ranked lists $\mathcal{R}$.

We also consider the problem of distance fusion, in which there are two or more image descriptors available. In this case, we define a function ρ_{rf} which takes as input the different sets of ranked lists $\{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_d\}$ computed by different descriptors, where d denotes the number of considered descriptors, and produces a distance score.

4. RECIPROCAL KNN DISTANCE LEARNING

The objective of *Reciprocal kNN Distance* is to define a more effective distance between two images by analyzing the reciprocal references among images at top positions of their ranked lists. This objective is mainly supported by the "*cluster hypothesis*", which states that closely related documents tend to be relevant to the same request [42]. In other words, if two query images are similar, the images well ranked for these queries probably refer to each other at top positions of their own ranked lists.

The analysis of the cluster hypothesis by the proposed *Reciprocal kNN Distance* is performed based on two approaches that have been receiving great attention on unsupervised distance learning recently: (i) the *similarity between ranked lists*, and (ii) the *reciprocal neighborhood*.

The ranked lists define relationships not only between pairs of images (as distance functions), but also among all images found in a ranked list [32]. In this sense, the ranked lists represent, by itself, a description of images with regard to the whole dataset and a relevant source of contextual information. In addition, a set of similar images tends to appear at top positions of the high-effective ranked lists and, therefore, those lists can be considered as a reliable tool to compare images.

On the other hand, the k-reciprocal nearest neighborhood relationship is a much stronger indicator of similarity than the unidirectional nearest neighborhood [33]. In this way, the k-reciprocal nearest neighborhood mitigates the risk of false positives at top positions of ranked lists. Given two similar images, an image descriptor is expected to produce ranked lists which present reciprocal references at the beginning of their ranked lists. When an image does not refer to the other image at the top positions of its ranked list, this behavior indicates a low confidence in the similarity between them.

The *Reciprocal kNN Distance* combines these two concepts for distance learning, which represents the main novelty of this work. The proposed distance is computed by modeling the similarity between ranked lists in terms of the amount of reciprocal references.

4.1 Reciprocal Neighborhood

Given a query image img_q, we can define a neighborhood set that contains the k most similar images to img_q as $\mathcal{N}(q,k)$. For the k-nearest neighbor query, we have $|\mathcal{N}(q,k)| = k$, which is formally defined as follows:

$$\mathcal{N}(q,k) = \{\mathcal{S} \subseteq \mathcal{C}, |\mathcal{S}| = k \wedge \forall img_i \in \mathcal{S}, img_j \in \mathcal{C} - \mathcal{S} : \tau_q(i) < \tau_q(j)\}. \quad (1)$$

The nearest neighbor relationships are not symmetric [16, 33], since $img_i \in \mathcal{N}(q,k)$ does not imply $img_q \in \mathcal{N}(i,k)$. The set of k-reciprocal nearest neighbors of image img_q can be defined [33] as:

$$\mathcal{N}_r(q,k) = \{img_i \in \mathcal{N}(q,k) \wedge img_q \in \mathcal{N}(i,k)\}. \quad (2)$$

Based on the reciprocal neighborhood set $\mathcal{N}_r(q,k)$, we define a binary function $f_r : \mathcal{C} \times \mathcal{C} \rightarrow \{0,1\}$ which determines if two images img_q, $img_i \in \mathcal{C}$ are reciprocal neighbors:

$$f_r(q,i) = |\mathcal{N}_r(q,k) \cap \{img_i\}|. \quad (3)$$

The function f_r returns 1 if img_q and img_i are reciprocal neighbors, and 0 otherwise.

4.2 Reciprocal kNN Distance

Given two images img_q, $img_i \in \mathcal{C}$ and their respective ranked lists $\tau_q, \tau_i \in \mathcal{R}$, the *Reciprocal kNN Distance* between them is computed based on the number of reciprocal neighbors at top positions of ranked lists. In addition, for each pair of reciprocal neighbors, a weight is computed proportionally to their position in the ranked lists τ_q and τ_i. The motivation consists in considering more relevant the incidence of reciprocal neighbors at top positions of ranked lists. The score based on the number of reciprocal neighbors and its respectively weights are given by the function $n_r(q,i)$, defined as follows:

$$n_r(q,i) = \frac{\sum_{j \in \mathcal{N}(q,k)} \sum_{l \in \mathcal{N}(i,k)} f_r(j,l) \times w_r(q,j) \times w_r(i,l)}{k^4}, \quad (4)$$

While the function f_r determines if a pair of images (img_j, img_l) are reciprocal neighbors, the weight is computed based on position of these images in ranked lists τ_q and τ_i, according to the function w_r, defined as follows:

$$w_r(q,j) = k + 1 - \tau_q(j). \quad (5)$$

The value of w_r is linearly decreasing, ranging from k assigned to the first position to 1, at the kth position. Notice that the divisor k^4 in Equation 4 is defined considering the maximum value of reciprocal neighbors (k^2) and the maximum values of w_r.

The *Reciprocal kNN Distance* is defined as the inverse of the number of reciprocal neighbors n_r, as follows:

$$\rho_r(q,i) = \frac{1}{1 + n_r(q,i)}. \quad (6)$$

Finally, we introduce a parameter L aiming at reducing the computational efforts needed to compute the distance learning procedure. Since the top positions of ranked lists are expected to contain the most relevant images related to the query image, the distance learning can be performed considering only the beginning of the ranked lists without significant loss of effectiveness (as detailed discussion in Section 5.1). From the L position to the end, the ranked lists remain the same. Therefore, we redefine ρ_r as follows:

$$\rho_r(q,i) = \begin{cases} \frac{1}{1+n_r(q,i)}, & \textit{if } \tau_q(i) \leq L, \\ \tau_q(i), & \textit{otherwise.} \end{cases} \quad (7)$$

4.3 Reciprocal kNN Distance Fusion

Different image descriptors and their respective distance functions may focus on different aspects of the images, which are often complementary to each other [3]. In this way, we aim at exploiting the unsupervised learning procedure based on the Reciprocal kNN Distance to combine different distance measures.

The proposed distance fusion approach is divided into in two main steps: (i) first, the sets of ranked lists computed by different descriptors are combined into a single set $\mathcal{R}_f$ through an intermediary distance ρ_f; (ii) next, the set $\mathcal{R}_f$ is used by the conventional Reciprocal kNN Distance presented in the previous section aiming at computing a final distance ρ_{rf}.

A traditional challenge in fusion tasks is to estimate the quality of each descriptor being combined. The main novelty of our fusion approach is the use of the number of reciprocal neighgors score n_r as an unsupervised estimation of quality of the descriptor for a given image. The main idea consists in considering the score $n_r(q,q)$ for a single ranked list of an image img_q[1].

The motivation of this approach is based on the conjecture that high-effective ranked lists are expected to present

[1]Repeated pairs of images are not considered for the computation of n_r score in this case, and therefore the divisor is equal to $k^4/2$.

a high number of similar images, and therefore, reciprocal neighbors are found at their top positions. Given a descriptor $\mathcal{D}_j$ and its corresponding computed set of ranked lists $\mathcal{R}_j$, we aim at estimating the capability of the descriptor determine the distance between img_q and img_i. Thus, we propose a quality estimation score $e_j(q,i)$:

$$e_j(q,i) = (1 + n_r(q,q)) \times (1 + n_r(i,i)). \quad (8)$$

The intermediary distance is computed by a multiplicative approach that considers, for each descriptor, the position from which on images img_q and img_i become reciprocal neighbors $(max(\tau_{j_q}(i), \tau_{j_i}(q)))$ according to the set of ranked lists $\mathcal{R}_j$. The relevance of the position computed by each descriptor for the combined distance function is determined by the quality estimation score $e_j(q,i)$. The intermediary function ρ_f is defined as follows:

$$\rho_f(q,i) = \prod_{j=1}^{d} max(\tau_{j_q}(i), \tau_{j_i}(q))^{e_j(q,i)}. \quad (9)$$

Finally, the intermediary function ρ_f is used to compute the set $\mathcal{R}_f$ with the combined ranked lists which is submitted to the Reciprocal kNN Distance learning procedure as a single descriptor.

5. EXPERIMENTAL EVALUATION

This section presents a set of conducted experiments for assessing the effectiveness of the proposed method. We analyzed and compared our method under several aspects. Section 5.1 discusses the impact of parameter values. Sections 5.2, 5.3, and 5.4 present the experimental results for the proposed approach considering various shape, color, and texture descriptors respectively. Section 5.5 discusses the use of our method for combining different descriptors. Section 5.6 presents the experimental results for object retrieval tasks, while Section 5.7 presents the results for multimodal image retrieval tasks. Finally, experiments aiming at comparing our results to state-of-the-art related methods are presented in Section 5.8.

5.1 Impact of Parameters

The computation of *Reciprocal kNN Distance* considers only two parameters: *(i)* k: the size of the neighborhood set; and *(ii)* L: the position at which the ranked lists are considered in the distance learning procedure.

To evaluate the impact of different parameter settings on the effectiveness of the method and for determining the best parameters values, we conducted two experiments considering the MPEG-7 [21] dataset. The MPEG-7 [21] dataset is a well-known shape dataset, composed of 1400 shapes divided in 70 classes. The *Mean Average Precision* (MAP) was considered as effectiveness measure.

In the first experiment, we fixed the value of $L = 200$ and varied the parameter k in the interval $[5, 50]$, considering the recently proposed Articulation-Invariant Representation (AIR) [13] shape descriptor. Figure 1 illustrates the results of MAP scores for different values of k.

A quickly grow of retrieval scores can be observed for the beginning of the curve, with $5 \leq k \leq 15$. However, between $k = 15$ and $k = 20$, no further improvements are obtained and the observed MAP scores is approximately 97%. Therefore, we set the parameter value as $k = 20$ for other experiments.

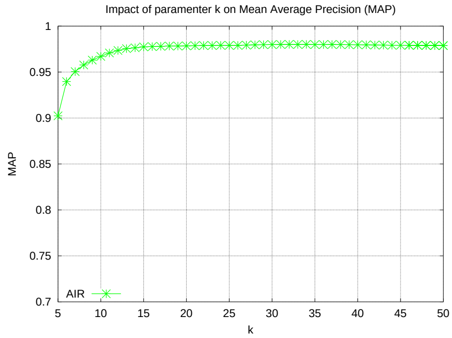

Figure 1: Impact of parameter k on MAP scores.

In the second experiment, we aim at measuring the impact of parameter L on effectiveness scores. As discussed before, the parameter L represents a trade-off between effectiveness and efficiency. In this, way evaluate three shape descriptors with different effectiveness: Contour Features Descriptor (CFD) [29], Aspect Shape Context (ASC) [25], and Articulation-Invariant Representation (AIR) [13]. We fixed the value of $k = 20$ varied the parameter L in the interval $[1, 1400]$. Figure 2 illustrates the variation of MAP score according to different values of L. Again, we can observe a quickly grow of retrieval scores in the beginning of the curve. As it can be observed, from $L = 400$ obtained MAP scores are maximum for all descriptors. We use $k = 20$ and $L = 400$ for most of our experimental evaluation.

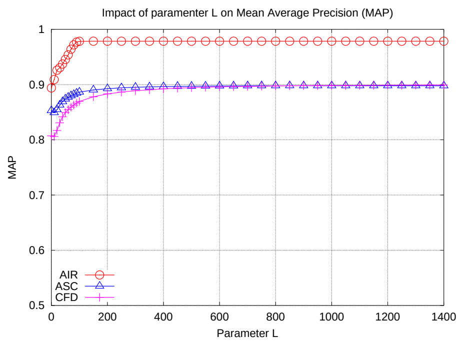

Figure 2: Impact of parameter L on MAP scores.

5.2 Shape Descriptors

We evaluate the use of our method for shape retrieval using the MPEG-7 [21] dataset, described in Section 5.1. Six shape descriptors were considered: Segment Saliences (SS) [9], Beam Angle Statistics (BAS) [1], Inner Distance Shape Context (IDSC) [24], Contour Features Descriptor (CFD) [29], Aspect Shape Context (ASC) [25], and Articulation-Invariant Representation (AIR) [13].

For evaluation, the so-called bull's eye score commonly used for this dataset was considered. 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 and is equivalent to Recall@40. Table 1 presents results considering the bull's eye score for shape descriptors on the MPEG-7 [21] dataset.

We also consider the more strict *accuracy* score on the MPEG-7 [21] dataset, which counts all matching objects within the 20 most similar candidates. Table 2 presents

Table 1: Reciprocal kNN Distance on the MPEG-7 [21] dataset considering the Bull's eye score.

Shape Descriptor	Bull's Eye Score	Reciprocal kNN Distance	Gain
SS [9]	43.99%	52.95%	+20.37%
BAS [1]	75.20%	80.74%	+7.37%
IDSC [24]	85.40%	90.75%	+6.26%
CFD [29]	84.43%	92.43%	+9.48%
ASC [25]	88.39%	93.05%	+5.27%
AIR [13]	93.67%	100%	+6.76%

Table 2: Reciprocal kNN Distance on the MPEG-7 [21] dataset considering the Accuracy score.

Shape Descriptor	Accuracy Score	Reciprocal kNN Distance	Gain
SS [9]	35.50%	43.48%	+22.47%
BAS [1]	67.33%	71.09%	+5.58%
IDSC [24]	77.21%	83.67%	+8.37%
CFD [29]	75.59%	85.94%	+13.69%
ASC [25]	80.66%	86.21%	+6.81%
AIR [13]	88.17%	94.05%	+6.67%

Table 3: Reciprocal kNN Distance for various retrieval tasks.

Descriptor	Type	Dataset	Score (MAP)	Reciprocal kNN Distance	Gain
SS [9]	Shape	MPEG-7	37.67%	46.55%	+23.57%
BAS [1]	Shape	MPEG-7	71.52%	75.59%	+5.69%
IDSC [24]	Shape	MPEG-7	81.70%	87.13%	+6.65%
CFD [29]	Shape	MPEG-7	80.71%	89.23%	+10.56%
ASC [25]	Shape	MPEG-7	85.28%	89.62%	+5.09%
AIR [13]	Shape	MPEG-7	89.39%	97.85%	+9.46%
GCH [37]	Color	Soccer	32.24%	34.86%	+8.13%
ACC [15]	Color	Soccer	37.23%	45.73%	+22.83%
BIC [36]	Color	Soccer	39.26%	46.32%	+17.98%
LBP [28]	Texture	Brodatz	48.40%	48.31%	-0.19%
CCOM [19]	Texture	Brodatz	57.57%	62.92%	+9.29%
LAS [38]	Texture	Brodatz	75.15%	77.46%	+3.07%

Table 4: Initial MAP scores for visual and textual retrieval on the UW dataset.

Descriptor	Type	Score (MAP)
GCH [37]	Visual - Color	31.75%
BIC [36]	Visual - Color	43.46%
JAC [44]	Visual - Color	**52.26%**
QCCH [14]	Visual - Texture	17.81%
LAS [38]	Visual - Texture	20.44%
HTD [46]	Visual - Texture	22.61%
DICE [23]	Textual	50.73%
OKAPI [35]	Textual	51.68%
COS [2]	Textual	41.80%
JACKARD [23]	Textual	50.29%
TF-IDF [2]	Textual	49.25%

the results for this measure. We can observe very significant gains in relation to the results observed for each descriptor initially, ranging from +5.27% to +20.37% for the bull's eye score and ranging from +5.58% to +22.47% for the accuracy measure. Notice that the accuracy gains are slightly greater, which indicates that the distance learning gains are mainly located on top positions of ranked lists.

We also evaluated the shape descriptors considering the MAP (*Mean Average Precision*) score. Results are presented in Table 3, along with the evaluation of other visual properties (color and texture). Positive gains can also be observed for all shape descriptors ranging from +5.09% to +23.57%.

5.3 Color Descriptors

We conducted experiments aiming at evaluating the Reciprocal kNN Distance for color descriptors. The experiments were conducted on a dataset [41] composed of images from 7 soccer teams, containing 40 images per class. Three color descriptors were considered: Border/Interior Pixel Classification (BIC) [36], Auto Color Correlograms (ACC) [15], and Global Color Histogram (GCH) [37]. Table 3 presents the experimental results considering MAP as score. We can observe a positive gain for all color descriptors ranging from +8.13% to +22.83%.

5.4 Texture Descriptors

The experiments considering texture descriptors were conducted on the Brodatz [5] dataset, a popular dataset for texture descriptors evaluation. The Brodatz [5] dataset is composed of 111 different textures. Each texture is divided into 16 blocks, such that 1,776 images are considered. We consider three well-known texture descriptors: Local Binary Patterns (LBP) [28], Color Co-Occurrence Matrix (CCOM) [19], and Local Activity Spectrum (LAS) [38]. Results considering MAP scores are presented in Table 3. We can observe positive gains ranging from +3.07% to +9.29%, except for LBP [28] descriptor, which presents a slightly loss. This case represents extreme situations, in which the visual descriptor completely confuse different classes of images and there are not enough information for the unsupervised learning approach.

5.5 Distance Fusion

We also evaluate the use of *Reciprocal kNN Distance* for distance fusion, aiming at combining different CBIR descriptors. We selected three shape descriptors with highest retrieval scores in distance learning tasks and evaluated the different combinations between them. Table 5 presents the fusion results. Besides MAP scores, we also present the accuracy and the bull's eye score on the MPEG-7 [21] dataset. Notice that the combination of CFD [29]+AIR [13] presents retrieval scores of 100% for the three considered measures, which means perfect retrieval results.

We also selected two color and texture descriptors, with the highest MAP scores in distance learning tasks. Table 6 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 color descriptors, for example, the fusion score achieves a MAP score of 47.40%, while the best descriptor in isolation yields only 39.26%.

5.6 Object Retrieval

We also evaluate the Reciprocal kNN Distance for ob-

Table 5: **Reciprocal kNN Distance for distance fusion on the MPEG-7 dataset.**

Descriptor	Bull's eye score	MAP	Accuracy
CFD [29]	84.43%	80.71%	75.59%
ASC [25]	88.39%	85.28%	80.66%
AIR [13]	93.67%	89.39%	88.17%
CFD+ASC	99.74%	99.18%	98.60%
CFD+AIR	100%	100%	100%
ASC+AIR	100%	99.99%	99.96%

ject retrieval tasks. The experiments were conducted on the ETH-80 [22] dataset, which is composed of 3,280 color images. Each image contains one single object, like tomatoes, cars, and cups, for example. The objects appear in many variations of rotation. For instance, a car was photographed from different angles. This dataset is equally divided into 8 classes where each class represents a different object, and all images have 128 × 128 pixels.

We evaluate our method considering four color descriptors: Border/Interior Pixel Classification (BIC) [36], Auto Color Correlograms (ACC) [15], Global Color Histogram (GCH) [37] and Color Structure Descriptor (CSD) [27]. Table 7 presents results considering the MAP scores. Positive gains were obtained for all considered descriptors ranging from +1.54% to +6.77%.

Table 7: Reciprocal kNN Distance for Object Retrieval on ETH-80 [22] dataset.

Descriptor	Score (MAP)	Reciprocal kNN Distance	Gain
BIC [36]	49.72%	53.08%	+6.76%
ACC [15]	48.50%	51.59%	+6.37%
CSD [27]	48.46%	51.74%	+6.77%
GCH [37]	41.62%	42.26%	+1.54%

5.7 Multimodal Retrieval

The UW dataset [11] was created at the University of Washington and consists of a roughly categorized collection of 1,109 images. The images include vacation pictures from various locations. These images are partly annotated using keywords. On the average, for each image the annotation contains 6 words. The maximum number of words per image is 22 and the minimum is 1. There are 18 categories, ranging from 22 images to 255 images per category.

The experiments consider eleven descriptors, which are listed below.

Visual Color Descriptors: we considered three color descriptors on experiments: Border/Interior Pixel Classification (BIC) [36], Global Color Histogram (GCH) [37] (both already used on Section 5.3), and the Joint Autocorrelogram (JAC) [44].

Visual Texture Descriptors: for texture we used the Homogeneous Texture Descriptor (HTD) [46], Quantized Compound Change Histogram (QCCH) [14], and Local Activity Spectrum (LAS) [38] (the last also considered in Section 5.4).

Textual Descriptors: five well-known text similarity measures are considered for textual retrieval, like the Cosine

Table 6: Reciprocal kNN Distance for distance fusion for color and texture descriptors.

Descriptor	Type	Dataset	Score (MAP)
ACC [15]	Color	Soccer	37.23%
BIC [36]	Color	Soccer	39.26%
BIC+ACC	Color	Soccer	47.40%
CCOM [19]	Texture	Brodatz	57.57%
LAS [38]	Texture	Brodatz	75.15%
LAS+CCOM	Texture	Brodatz	82.84%

similarity measure (COS), Term Frequency - Inverse Document Frequency (TF-IDF), and the Dice coefficient (DICE).

Table 4 presents the MAP scores for each descriptor isolated. Experiments were conducted considering different scenarios: using all descriptors of each modality; using only the best descriptors. Two baselines are also considered in the experiments: the traditional Borda [50] method and the recently proposed Reciprocal Rank Fusion [7]. Table 8 presents the MAP results the Reciprocal kNN Distance.

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 (74.75%) presents a very significant gain of +43.04% over the best individual descriptor in isolation.

Figure 3 presents example results obtained by visual and textual descriptors in isolation and using the Reciprocal kNN Distance in distance fusion tasks.The gains can also be observed in Figure 4, which illustrates the Precision × Recall curve for the considered descriptors and for the Reciprocal kNN Distance.

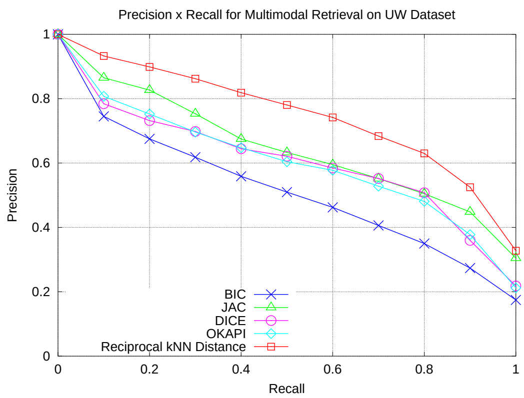

Figure 4: Precision × Recall curve for multimodal retrieval on UW dataset [11].

5.8 Comparison with Other Approaches

Finally, we also evaluate our method in comparison with other state-of-the-art post-processing methods. We use the MPEG-7 [21] dataset commonly used for post-processing methods evaluation and comparison. We first considered the bull's eye score, often used for comparisons. Table 9 presents results of the proposed *Reciprocal kNN Distance* in comparison with several other post-processing methods recently proposed in the literature. We report the two best results of our approach and for each of most recent methods. Note that the results for distance learning and distance fusion presents better effectiveness performance when com-

Table 8: Reciprocal kNN Distance on multimodal retrieval tasks (*MAP* as score).

Retrieval Task	Descriptors	Reciprocal kNN Distance	Borda [50]	Reciprocal Fusion [7]
Visual	All visual descriptors	47.91%	40.29%	43.29%
Textual	All textual descriptors	63.74%	53.07%	53.14%
Multimodal	All descriptors	70.09%	54.89%	59.34%
Visual	BIC [36]+JAC [44]	74.05%	52.54%	53.00%
Textual	DICE [23]+OKAPI [35]	64.35%	54.57%	54.31%
Multimodal	BIC [36]+JAC [44]+DICE [23]+OKAPI [35]	**74.75%**	61.91%	63.67%

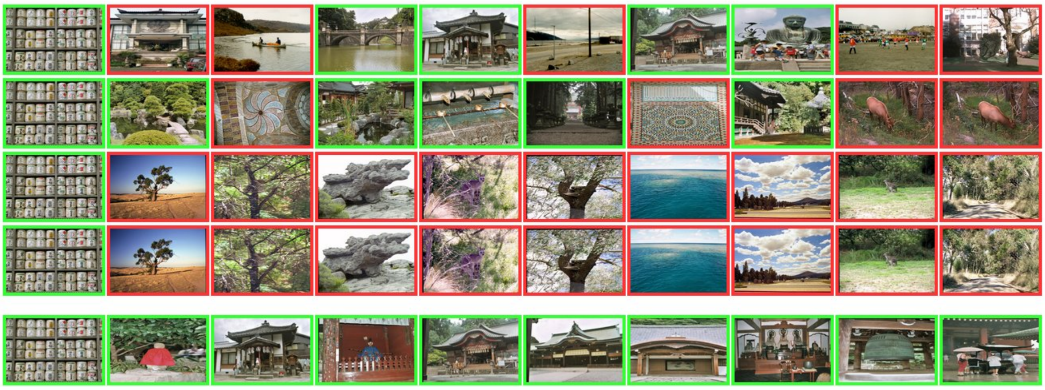

Figure 3: Example of results for a multimodal image retrieval task considering a query image of class "Japan" (first column). Each line presents the retrieved images by visual and textual descriptors (BIC [36], JAC [44], DICE [23], and OKAPI [35] respectively), with green and red borders for relevant and non-relevant images. The last line presents the results of the proposed Reciprocal kNN Distance.

pared to various recently proposed methods. The *Reciprocal kNN Distance* achieves a bull's eye score of 100% for the AIR [13] shape descriptor.

We also considered the accuracy score, a more strict measure used recently due to the saturation of the bull's eye score. Table 10 presents the results for the accuracy measure. Note that the Reciprocal kNN Distance applied to the combination of only two descriptors CFD [29] + AIR [13] reached a perfect retrieval score, obtained by other state-of-the-art method only combining three descriptors.

6. CONCLUSIONS

In this work, we have presented a novel unsupervised learning approach called *Reciprocal kNN Distance* aiming at improving the effectiveness of distance measures on image retrieval tasks. The main idea consists in computing the similarity among ranked lists by analyzing the reciprocal references at top positions of ranked lists.

A large set of experiments was conducted considering different descriptors and datasets. Experimental results demonstrated the applicability of our method on different scenarios, considering visual image retrieval (shape, color, and texture descriptors), object retrieval and multimodal retrieval (visual and textual descriptors). In addition, the proposed approach also achieves very high effectiveness performance when compared with recent state-of-the-art methods on well-known datasets.

Future work focuses on: (*i*) using our distance fusion approach for combining local and global descriptors; and (*ii*) the implementation of the proposed distance learning approach by considering parallel architectures.

7. ACKNOWLEDGMENTS

The authors are grateful to São Paulo Research Foundation - FAPESP (grant 2013/08645-0), CNPq (grants 306580/2012-8 and 484254/2012-0), CAPES, AMD, and Microsoft Research for the financial support.

8. REFERENCES

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

[2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman, 1999.

[3] X. Bai, B. Wang, X. Wang, W. Liu, and Z. Tu. Co-transduction for shape retrieval. In *ECCV*, volume 3, pages 328–341, 2010.

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

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

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

[7] G. V. Cormack, C. L. A. Clarke, and S. Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *ACM SIGIR*, pages 758–759, 2009.

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

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

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

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

[12] Z. Ghahramani. Unsupervised Learning. In *Advanced Lectures on Machine Learning*, pages 72–112. 2004.

Table 9: Post-processing methods comparison on the MPEG-7 dataset - *Bull's eye score.*

Shape Descriptors		
DDGM [40]	-	80.03%
CFD [29]	-	84.43%
IDSC [24]	-	85.40%
SC [4]	-	86.80%
ASC [25]	-	88.39%
AIR [13]	-	93.67%
Post-Processing Methods		
Algorithm	**Descriptor(s)**	**Bull's eye score**
Graph Transduction [47]	IDSC	91.00%
Locally C. Diffusion Process [48]	IDSC	93.32%
Mutual kNN Graph [18]	IDSC	93.40%
Locally C. Diffusion Process [48]	ASC	95.96%
Pairwise Recommendation [31]	CFD	96.15%
Tensor Product Graph [49]	ASC	96.47%
Co-Transduction [3]	SC+DDGM	97.45%
Self-Smoothing Operator [17]	SC+IDSC	97.64%
Co-Transduction [3]	SC+IDSC	97.72%
Self-Smoothing Operator [17]	SC+IDSC+DDGM	99.20%
Pairwise Recommendation [31]	CFD+IDSC	99.52%
RL-Sim [32]	CFD+ASC	99.65%
Reciprocal kNN Distance	**CFD+ASC**	**99.74%**
RL-Sim [32]	AIR	99.94%
Tensor Product Graph [49]	AIR	99.99%
Reciprocal kNN Distance	**AIR**	**100%**

Table 10: Post-processing methods comparison on the MPEG-7 dataset - *Accuracy score.*

Post-Processing Methods		
Algorithm	**Descriptor(s)**	**Accuracy**
Co-Transduction [3]	IDSC+DDGM	95.12%
Co-Transduction [3]	SC+IDSC+DDGM	95.24%
Cross Diffusion Process [43]	SC+IDSC	99.86%
Reciprocal kNN Distance	**ASC+AIR**	**99.96%**
Cross Diffusion Process [43]	SC+IDSC+DDGM	100%
Reciprocal kNN Distance	**CFD+AIR**	**100%**

[13] R. Gopalan, P. Turaga, and R. Chellappa. Articulation-invariant representation of non-planar shapes. In *ECCV'2010*, volume 3, pages 286–299, 2010.

[14] C.-B. Huang and Q. Liu. An orientation independent texture descriptor for image retrieval. In *ICCCAS 2007*, pages 772–776, 2007.

[15] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Image indexing using color correlograms. In *CVPR*, pages 762–768, 1997.

[16] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek. Accurate image search using the contextual dissimilarity measure. *PAMI*, 32(1):2–11, 2010.

[17] J. Jiang, B. Wang, and Z. Tu. Unsupervised metric learning by self-smoothing operator. In *ICCV*, pages 794–801, 2011.

[18] P. Kontschieder, M. Donoser, and H. Bischof. Beyond pairwise shape similarity analysis. In *ACCV*, pages 655–666, 2009.

[19] V. Kovalev and S. Volmer. Color co-occurence descriptors for querying-by-example. In *International Conference on Multimedia Modeling*, page 32, 1998.

[20] S. Lafon and A. B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *PAMI*, 28(9):1393–1403, 2006.

[21] L. J. Latecki, R. Lakmper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, pages 424–429, 2000.

[22] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, volume 2, pages II–409–15 vol.2, 2003.

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

[24] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *PAMI*, 29(2):286–299, 2007.

[25] H. Ling, X. Yang, and L. J. Latecki. Balancing deformability and discriminability for shape matching. In *ECCV*, volume 3, pages 411–424, 2010.

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

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

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

[29] D. C. G. Pedronette and R. da S. Torres. Shape retrieval using contour features and distance optmization. In *VISAPP*, volume 1, pages 197 – 202, 2010.

[30] D. C. G. Pedronette and R. da S. Torres. Exploiting clustering approaches for image re-ranking. *Journal of Visual Languages and Computing*, 22(6):453–466, 2011.

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

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

[33] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR*, pages 777 –784, june 2011.

[34] T. Qin, T.-Y. Liu, X.-D. Zhang, D.-S. Wang, and H. Li. Global ranking using continuous conditional random fields. In *NIPS*, pages 1281–1288, 2008.

[35] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In *TREC*, pages 109–126, 1994.

[36] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM*, pages 102–109, 2002.

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

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

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

[40] Z. Tu and A. L. Yuille. Shape matching and recognition - using generative models and informative features. In *ECCV*, pages 195–209, 2004.

[41] J. van de Weijer and C. Schmid. Coloring local feature extraction. In *ECCV*, pages 334–348, 2006.

[42] C. J. van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, London, 1979.

[43] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, and Z. Tu. Unsupervised metric fusion by cross diffusion. In *CVPR*, pages 3013 –3020, 2012.

[44] A. Williams and P. Yoon. Content-based image retrieval using joint correlograms. *MTAP*, 34(2):239–248, August 2007.

[45] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *ACCV'09*, pages 88–97, 2009.

[46] P. Wu, B. S. Manjunanth, S. D. Newsam, and H. D. Shin. A texture descriptor for image retrieval and browsing. In *CBAIVL*, pages 3–7, 1999.

[47] X. Yang, X. Bai, L. J. Latecki, and Z. Tu. Improving shape retrieval by learning graph transduction. In *ECCV*, volume 4, pages 788–801, 2008.

[48] X. Yang, S. Koknar-Tezel, and L. J. Latecki. Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *CVPR*, pages 357–364, 2009.

[49] X. Yang, L. Prasad, and L. Latecki. Affinity learning with diffusion on tensor product graph. *PAMI*, PP(99):1, 2012.

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

[51] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas. Query specific fusion for image retrieval. In *ECCV*, pages 660–673, 2012.