Unsupervised measures for estimating the effectiveness of image retrieval systems

Daniel Carlos Guimarães Pedronette
Department of Statistics, Applied Mathematics and Computing
State University of São Paulo (UNESP) - Rio Claro - Brazil
daniel@rc.unesp.br

Ricardo da S. Torres
Recod Lab - Institute of Computing
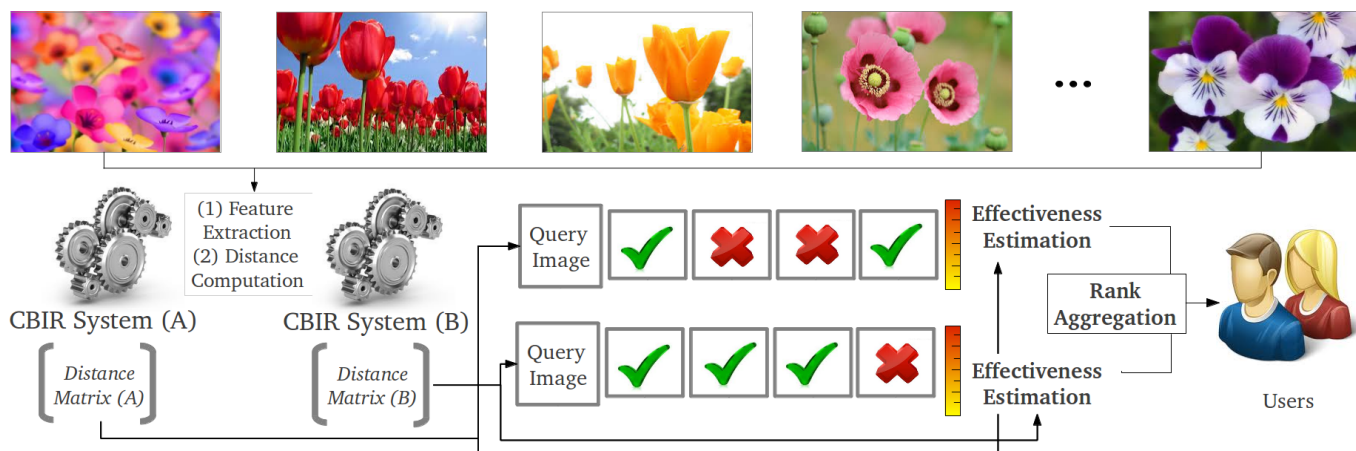University of Campinas (UNICAMP) - Campinas - Brazil
rtorres@ic.unicamp.br

Fig. 1. Generic workflow for the use of unsupervised effectiveness estimation measures in Content-Based Image Retrieval tasks.

Abstract—The main objective of Content-Based Image Retrieval (CBIR) systems is to retrieve a ranked list containing the most similar images of a collection given a query image, by taking into account their visual content. Although these systems represent a very promising approach, in many situations is very challenging to assure the quality of returned ranked lists. Supervised approaches rely on training data and information obtained from user interactions to identify and then improve low-quality results. However, these approaches require a lot of human efforts which can be infeasible for many systems.

In this paper, we present two novel unsupervised measures for estimating the effectiveness of ranked lists in CBIR tasks. Given an estimation of the effectiveness of ranked lists, many CBIR systems can, for example, emulate the training process, but now without any user intervention. Improvements can also be achieved on several unsupervised approaches, such as re-ranking and rank aggregation methods, once the estimation measures can help to consider more relevant information by distinguishing effective from non-effective ranked lists. Both proposed measures are computed using a novel image representation of ranked lists and distances among images considering a given dataset. The objective is to exploit the visual patterns encoded in the image representations for estimating the effectiveness of ranked lists. Experiments involving shape, color, and texture descriptors demonstrate that the proposed approaches can provide accurate estimations of the quality in terms of effectiveness of ranked lists. The use of proposed measures are also evaluated in image retrieval tasks aiming at improving the effectiveness of rank aggregation approaches.

Keywords-effectiveness estimation, content-based image retrieval, rank aggregation

I. INTRODUCTION

A large set of technological advances in image acquisition, storage, and sharing have contributed to the huge growth of image collections. A very promising approach for dealing with all available visual data relies on the use of Content-Based Image Retrieval (CBIR) systems. Content-Based Image Retrieval can be broadly defined as any technology that in principle helps to organize digital picture archives by their visual content [1]. A common application of these systems consists in supporting image searches. The main objective of Content-based Image Retrieval (CBIR) systems is to retrieve a list containing the most similar images in a collection given a query image, according to their visual properties (such as, shape, color, and texture). In the following, collection images are ranked in decreasing order of similarity, according to a given image descriptor.

An image content descriptor is characterized by [2]: *(i)* an extraction algorithm that encodes image features into feature vectors; and *(ii)* a similarity measure used to compare two images. The similarity between two images is computed as a function of the distance of their feature vectors. A common limitation of several approaches is the well-known *semantic*

gap between low-level features and higher-level concepts, which basically consists in the use of visual similarity for inferring semantic similarity. Motivated by these limitations, many supervised learning approaches have been proposed. Relevance Feedback methods [3], [4], for example, were incorporated into CBIR systems with the aim of exploiting interactions for learning users needs. Although very effective, these approaches require a lot of human efforts for obtaining enough training data, which can be infeasible for many real-world systems.

Aiming at overcoming these problems, efforts were put on unsupervised approaches. Methods have been proposed for analyzing the relations among all images in a given collection, without the need of user intervention [5]–[12]. Contextual information has been exploited to improve the effectiveness of CBIR systems. More specifically, contextual information refers to information encoded in the ranked lists defined by a CBIR system [13], which is used to update similarity scores. Broadly, the general objective of these unsupervised methods is somehow mimic the human behavior on judging the similarity among objects by considering specific *contexts* [14].

Contextual information can also be exploited for unsupervised estimation of effectiveness of CBIR systems. Given an accurate estimation of the effectiveness of ranked lists, many CBIR systems can, for example, emulate training procedures, but now without any user intervention. Improvements can also be achieved on several unsupervised approaches, such as re-ranking and rank aggregation methods, once the estimation measures can help to consider more relevant information as it allows for distinguishing effective from non-effective ranked lists.

In this paper, we present two novel unsupervised measures aiming at approximating the human judge of similarity by estimating the effectiveness measures of ranked lists. Figure 1 illustrates an example of the use of unsupervised effectiveness estimation measures. The main idea consists in exploiting contextual information encoded in distances among images for estimating the effectiveness of ranked lists. Given an effectiveness estimation of ranked lists, final results provided to the users can be improved. Both measures proposed in this paper are computed using a color image representation, named neighborhood image, based on information extracted from both ranked lists and distances among images. As discussed along the paper, the proposed image representations provide a rich contextual information related to the dataset and the relationships among images.

We conducted an experimental evaluation by computing the correlation of proposed measures with established effectiveness measures. We compared our proposed measures with a recently proposed approach considering shape, color, and texture descriptors on different datasets. We also evaluated the application of proposed measures in rank aggregation methods. Experimental results demonstrate that the proposed approaches can provide accurate estimations of the effectiveness of ranked lists.

II. IMAGE REPRESENTATION OF NEIGHBORHOOD SIMILARITY

Let $\mathcal{C}=\{img_1, img_2, \ldots, img_n\}$ be an *image collection*. Let $\mathcal{D}$ be an *image descriptor* which defines a distance function between two images img_i and img_j as $\rho(img_i, img_j)$. For readability purposes, we use the notation $\rho(i,j)$ for denoting the distance between images img_i and img_j.

Based on the distance function ρ, a ranked list τ_q can be computed in response to a given query image img_q. The ranked list $\tau_q=(img_1, img_2, \ldots, img_n)$ can be defined as a permutation of the subset $\mathcal{C}_s \subset \mathcal{C}$, which contains the most similar images to query image img_q, such that $|\mathcal{C}| = n$. A permutation τ_q is a bijection from the set $\mathcal{C}_s$ 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.

Our goal is to create an image representation of the neighborhood similarity of a given image, considering distances and ranked lists computed by CBIR descriptors. The image representation approach provides a visual representation of the neighborhood similarity and enables the use of image processing techniques, a well-established area. We propose a novel color image representation, inspired by a recently proposed gray scale image representation [15], [16] used by re-ranking an rank aggregation algorithms. As discussed in next section, information encoded in the image representation can be exploited for computing the unsupervised measures for effectiveness estimation.

We can formally define the proposed image representation as follows. Let the *neighborhood image* $\hat{I}$ be a color image defined by the pairs (D_I, r), (D_I, g), and (D_I, b), where D_I is a finite set of pixels (points in $\mathbb{N}^2$, defined by a pair (x, y)) and $r, g, b : D_I \to \mathbb{N}$ are functions that assign to each pixel $p \in D_I$ a natural number for each color channel red, green, and blue. We define the values of functions in terms of the distance function ρ and the ranked list τ_q of the query image img_q.

Let $\tau_q=(img_1, img_2, \ldots, img_n)$ be the ranked list computed by the image descriptor for the query image img_q. The axis of *neighborhood image* $\hat{I}$ are ordered according to the order defined by ranked list τ_q. Let $p(x, y)$ be the pixel at the position (x, y) and let img_i, img_j be images in the ranked list τ_q, such that the position of img_i is given by $\tau_q(i) = x$ and $\tau_q(j) = y$.

The value of $r(x, y)$ (function that defines the red scale of pixel $p(x, y)$) is defined as the distance between img_i and img_j as follows: $r(x, y) = \bar{\rho}(i, j)$, where $\bar{\rho}$ is defined by the distance function ρ normalized in the interval [0,255]. The values returned by functions $g(x, y)$ and $b(x, y)$ are defined as the distances to img_i and img_j according to the query image img_q, that is, as the difference between distances from img_q to img_i and img_j. The value of these functions are computed as follows: $g(x, y) = b(x, y) = |\bar{\rho}(q, i) - \bar{\rho}(q, j)|$. Figure 6 illustrates how the *neighborhood image* is constructed.

Our hypothesis is that the *neighborhood image* can provide
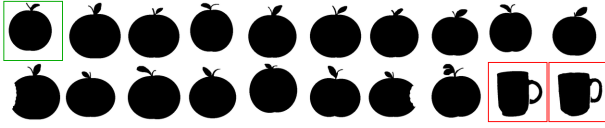
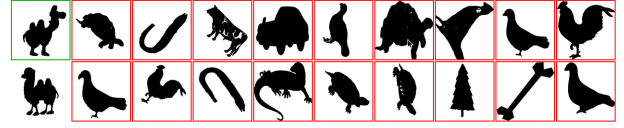Fig. 2.   High-effective ranked list.

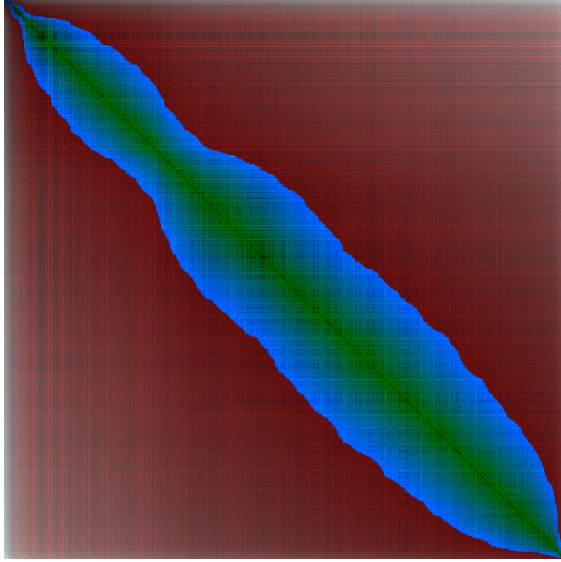Fig. 3.   Low-effective ranked list.

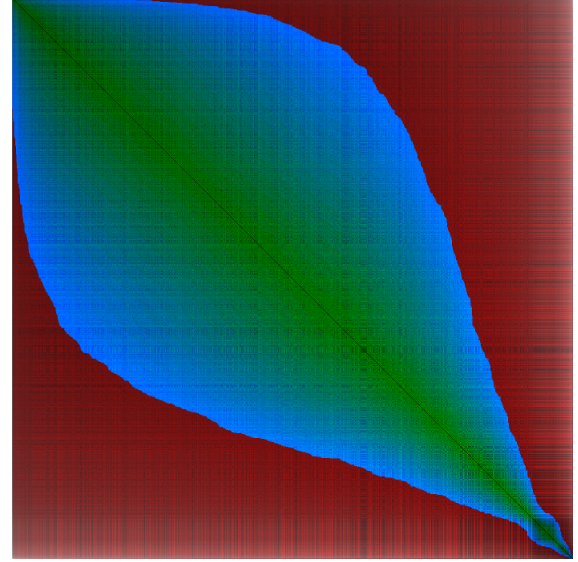Fig. 4.   Image representation for high-effective ranked list.

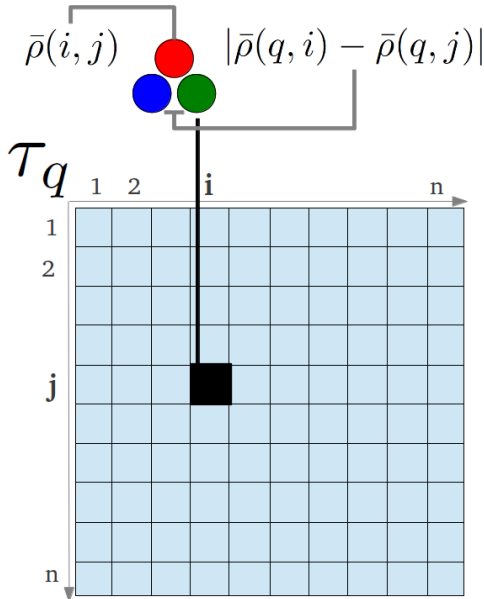Fig. 5.   Image representation for low-effective ranked list.

Fig. 6.   Computation of RGB channels of neighborhood image.

useful information about the effectiveness of the ranked list τ_q. An example of an effective ranked list (from MPEG-7 [17] collection) is illustrated in Figure 2. The first image (with green border) represents the query image and remaining images the computed ranked list by the CFD [18] shape descriptor. The non-similar images to the query image are presented with red borders. The respective *neighborhood image* representing this ranked list is illustrated in Figure 4. An analogous example for a non-effective ranked list is illustrated in Figures 3 and 5.

As we can observe, the two *neighborhood images* present very dissimilar visual patterns. Low distance values (similar images) are associated with red dark pixels in the image, while high values (non-similar images) refers to red bright pixels. Considering an effective ranked list, a *dark region* in shades of red is produced in the top left corner of the *neighborhood image* specially around the main diagonal (as we can observe in Figure 4).

The other visual pattern that differs the two images (from effective and non-effective ranked lists) is concerning with the shades of blue and green. Shades of blue and green represent the difference between distances from img_q to img_i and img_j. Small differences between distances $\rho(q,i)$ and $\rho(q,j)$ indicate low confidence of the image descriptor to differ img_i from img_j concerning with their similarity to img_q.

For non-effective ranked lists, the combination of bright shades of red and dark shades of blue and green produces a region with shades of blue and green in the top left corner of the image (as we can observe in Figure 5). Our objective is to exploit these visual patterns for estimating the effectiveness of ranked lists, as described in next section.

III. Effectiveness Estimation Measures

In this section, we present two unsupervised effectiveness estimation measures based on the representation defined by the *neighborhood image*. For both measures, we consider the top-left corner of the image, with size of $k \times k$. This region is related to the beginning of ranked lists, where the precision of results is higher than any other region of the image.

A. Neighborhood Distance Measure (NDM)

This Neighborhood Distance Measure (NDM) is computed based on occurrence of dark pixels in the red channel at the top-left corner of the *neighborhood image*. These pixels indicate a low distance among images at first positions of the ranked lists being analysed. The NDM measure is computed as follows:

$$NDM = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \frac{1}{1 + \sqrt{i^2 + j^2} \times |i - j|} \times t(i,j), \quad (1)$$

where $t(i,j)$ represents a threshold function for identifying dark pixels, so that $t(i,j) = 1$ if $r(i,j) < l$ and $t(i,j) = 0$ otherwise. The threshold l is computed based on average and maximum distance values contained in $k \times k$ square in top left corner:

$$l = \frac{avg(\rho(i,j))}{max(\rho(i,j))} \times 255 \quad (2)$$

with $i, j < k$. The dividend terms aim at assigning higher weights to pixels at the beginning of ranked lists and close to the main diagonal.

B. Neighborhood Distance Variation Measure (NDVM)

The Neighborhood Distance Variance Measure (NVM) is computed based on the occurrence of high values of blue and green channels at the top left corner of image. These pixels indicate the capacity of the image descriptor for discriminating images at the first positions of the ranked lists. In this way, the greater the variation of distances, the greater are the values of blue and green channels and higher the expected effectiveness of the ranked list. The NVDM measure is computed as follows:

$$NVDM = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} (g(i,j) + b(i,j)) \times (\sqrt{i^2 + j^2} \times |i - j|). \quad (3)$$

As for NDM measure, higher weights are assigned to pixels at the beginning of ranked lists and close to the main diagonal, which present higher precision.

IV. Experimental Evaluation

This section presents conducted experiments for demonstrating the applicability of proposed measures. We analyzed the measures under several aspects, considering different descriptors and datasets. We also introduce the use of the proposed measures in rank aggregation tasks.

A. Datasets and Descriptors

Three datasets and eleven CBIR descriptors (five shape descriptors, three color descriptors, and three texture descriptors) are considered. We briefly describe the datasets and descriptors in the following.

B. Shape

We evaluate the use of our method with five shape descriptors: Segment Saliences (SS) [19], Beam Angle Statistics (BAS) [20], Inner Distance Shape Context (IDSC) [21], Contour Features Descriptor (CFD) [18], and Aspect Shape Context (ASC) [22].

The experiments with shape descriptors were conducted on the MPEG-7 dataset. The MPEG-7 [17] dataset is a well-known shape collection, commonly used for shape descriptors and post-processing methods evaluation and comparison. It is composed of 1400 shapes divided into 70 classes of 20 images each. The size of images range from (50×48) to (526×408) pixels. Figure 7 presents some examples of images of the MPEG-7 dataset.
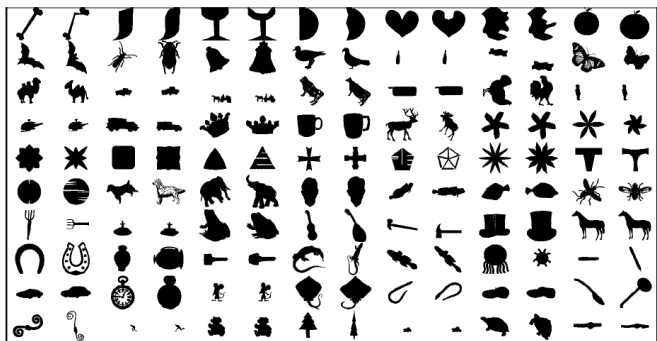

Fig. 7. Examples of shapes in the MPEG-7 dataset.

C. Color

We evaluate our method with three color descriptors: Border/Interior Pixel Classification (BIC) [23], Auto Color Correlograms (ACC) [24], and Global Color Histogram (GCH) [25]. The experiments were conducted on a dataset used in [26] and composed of images from 7 soccer teams, containing 40 images per class. The size of images range from (198×148) to (537×672) pixels. Some samples of this dataset are illustrated in Figure 8.

Fig. 8. Examples of images in the Soccer dataset [26].

D. Texture

The experiments consider three well-known texture descriptors: Local Binary Patterns (LBP) [27], Color Co-Occurrence Matrix (CCOM) [28], and Local Activity Spectrum (LAS) [29]. We used the Brodatz [30] dataset, a popular dataset for texture descriptors evaluation was considered. The Brodatz dataset is composed of 111 different textures of size (512×512) pixels. Each texture is divided into 16 blocks (128×128) pixels of non-overlapping sub images, such that 1776 images are considered. Some examples of textures are presented in Figure 9.
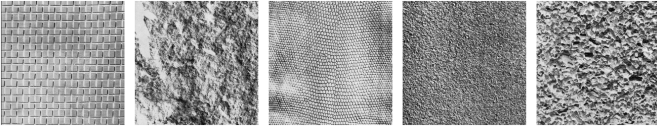

Fig. 9. Examples of Brodatz [30] texture images.

E. Effectiveness Correlation Analysis

This section aims at evaluating the use of the proposed measures as an unsupervised estimation of the effectiveness of ranked lists. We analyzed the correlation between the proposed measures and the well-known average precision measure, considering shape, color, and texture descriptors and dataset. The objective of the correlation analysis is to assess if the proposed measures are correlated with a traditional and widely used CBIR effectiveness score (MAP). If the proposed measures are correlated with MAP, that means they can be used to estimate the quality of ranked list in an unsupervised way.

We also compared the measures presented in this paper with a recently proposed cohesion [10] measure. The cohesion measure is used in re-ranking algorithms to estimate the quality of ranked lists and to define convergence criterion[1].

1) Experimental Protocol: We compared the proposed measures with the Average Precision (AP), which is is one of the most used effectiveness measure. Let q be a query item and let N_r be the number of relevant items in a collection for a given query q. Let $\langle r_i | i = 1, 2, \ldots, d \rangle$ be a ranked relevance vector to depth d, where r_i indicates the relevance of the ith ranked document scored as either 0 (not relevant) or 1 (relevant), the AP is defined as follows:

[1]We used k=8 for cohesion measure, according to [10].

$$AP = \frac{1}{N_r} \sum_{i=1}^{d} \left(\frac{r_i}{i} \sum_{j=1}^{i} r_j \right). \tag{4}$$

We use an statistical measure to evaluate the magnitude of a relationship among the proposed measures and the average precision. We analyze this relationship by using Pearson's Correlation Coefficient, given by:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}. \tag{5}$$

Pearson's correlation coefficient r for continuous data ranges from -1 to +1, where $r = 1$ indicates a perfect positive linear relationship and $r = -1$ a perfect decreasing linear relationship The closer the coefficient is to 1, the stronger the correlation between the variables.

2) Impact of Parameters: The computation of the proposed measures considers the parameter k, which represents the number of neighbors considered for the effectiveness estimation. To evaluate the influence of this parameter and for determining the best value of k, we conducted a set of experiments. We use the MPEG-7 [17] dataset and the CFD [18] shape descriptor. The Pearson correlation between the proposed measures and the average precision are computed ranging k in the interval $[5, 100]$ with variations of 5.

Figure 12 illustrates the results of correlation score for different values of k. We observed that best correlation score converged for values $k = 35$ for NDM and $k = 50$ for NVDM. We used these values in all experiments.
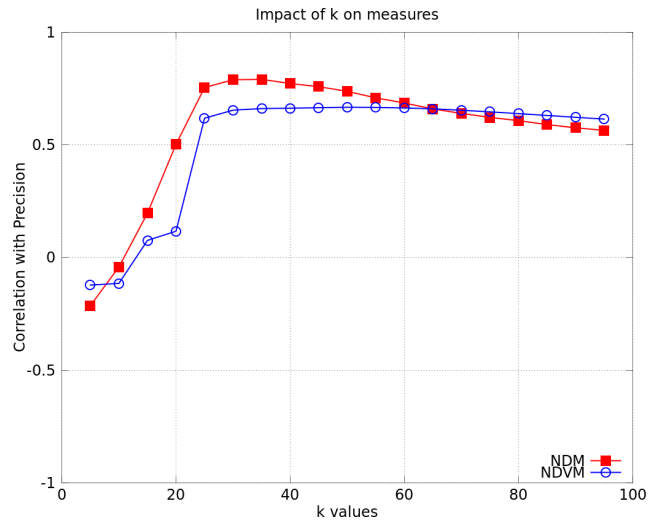

Fig. 12. Impact of parameters: different values of k.

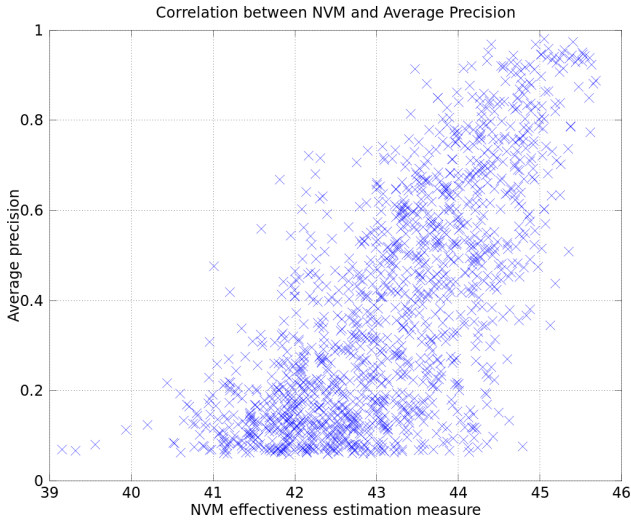| Descriptor	Type	Dataset	MAP	NDM	NVDM	Cohesion [10]
SS [19]	Shape	MPEG-7	37.67%	0.71	**0.81**	0.80
BAS [20]	Shape	MPEG-7	71.52%	**0.79**	0.76	0.53
CFD [18]	Shape	MPEG-7	80.71%	**0.79**	0.67	0.50
IDSC [21]	Shape	MPEG-7	81.70%	**0.76**	0.60	0.36
ASC [22]	Shape	MPEG-7	85.28%	**0.75**	0.58	0.36
GCH [25]	Color	Soccer	32.24%	0.13	**0.26**	0.15
ACC [24]	Color	Soccer	37.23%	0.28	**0.49**	0.34
BIC [23]	Color	Soccer	39.26%	0.23	0.44	**0.46**
LBP [27]	Texture	Brodatz	48.40%	0.45	**0.54**	**0.54**
CCOM [28]	Texture	Brodatz	57.57%	0.08	**0.71**	0.68
LAS [29]	Texture	Brodatz	75.15%	0.67	**0.71**	0.68
Average				0.51	**0.60**	0.49

Fig. 10. Example of correlation between the NDM measure and average precision.
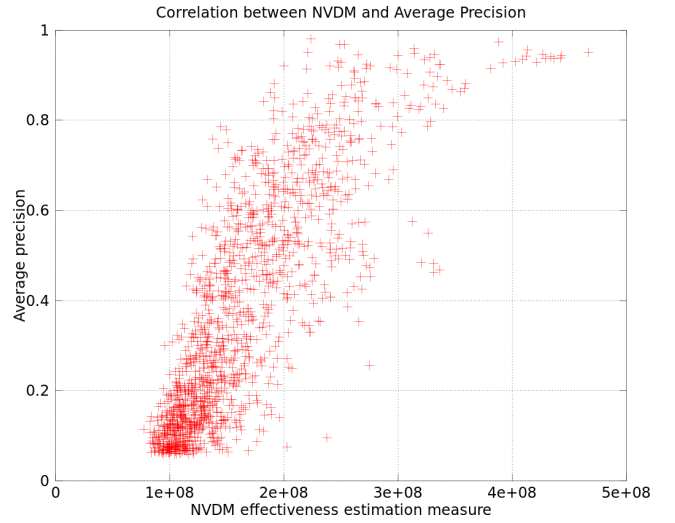

Fig. 11. Example of correlation between the NVDM measure and average precision

3) Correlation Results: Table I shows the obtained results for the correlation analysis. For each pair dataset-descriptor, we report the MAP (Mean Average Precision) obtained by the descriptor and correlation between effectiveness estimation measures and the average precision. For each descriptor, the highest correlation is presented in bold. We can observe that the NDM measure presents the best correlation scores for the MPEG-7 [17] dataset (which presents higher MAP scores), reaching 0.79 for some descriptors.

The NVDM measure presents high correlation scores for most of descriptors, including low and high MAP scores. The NVDM measure also overcomes the cohesion [10] measure for almost all descriptors, achieving an average correlation score of 0.6 and the highest score of 0.81 for the SS [19] descriptor.

Figures 10 and 11 illustrate examples of correlation between effectiveness estimation measures and average precision. Both examples considered the MPEG-7 [17] dataset and the SS [19] shape descriptor. Figure 10 presents results for the NVM mea-sure, while Figure 11 shows results for the NVDM measure. Each point in the graph represents a collection image. We can observe that the graphs approximate a linear relationship with a positive slope.

F. Rank Aggregation Methods

A direct application of our proposed effectiveness estimation measures consists in their use on rank aggregation methods. Different CBIR descriptors produce different results. Further, it is intuitive that different descriptors may provide different but complementary information. Basically, rank aggregation methods aim at combining different ranked lists in order to obtain a more accurate one.

However, rank aggregation approaches operate on unsupervised way, without training data. In this way, our estimation measures can be used for improving the effectiveness of combined ranked lists. In our experimental evaluation, we consider two methods: the traditional Borda [31] and the

Descriptor	Type	Dataset	Method	Measure	Score (MAP)
CFD [18]+ASC [22]	Shape	MPEG-7	Borda	-	91.12%
CFD [18]+ASC [22]	Shape	MPEG-7	Borda	NDM	91.21%
CFD [18]+ASC [22]	Shape	MPEG-7	Borda	NVDM	91.83%
CFD [18]+ASC [22]	Shape	MPEG-7	Reciprocal	-	93.80%
CFD [18]+ASC [22]	Shape	MPEG-7	Reciprocal	NDM	**94.02%**
CFD [18]+ASC [22]	Shape	MPEG-7	Reciprocal	NVDM	93.80%
BIC [23]+ACC [24]	Color	Soccer	Borda	-	38.81%
BIC [23]+ACC [24]	Color	Soccer	Borda	NDM	38.81%
BIC [23]+ACC [24]	Color	Soccer	Borda	NVDM	38.86%
BIC [23]+ACC [24]	Color	Soccer	Reciprocal	-	38.88%
BIC [23]+ACC [24]	Color	Soccer	Reciprocal	NDM	38.89%
BIC [23]+ACC [24]	Color	Soccer	Reciprocal	NVDM	**38.94%**
LAS [29]+CCOM [28]	Texture	Brodatz	Borda	-	73.92%
LAS [29]+CCOM [28]	Texture	Brodatz	Borda	NDM	73.92%
LAS [29]+CCOM [28]	Texture	Brodatz	Borda	NVDM	74.49%
LAS [29]+CCOM [28]	Texture	Brodatz	Reciprocal	-	75.49%
LAS [29]+CCOM [28]	Texture	Brodatz	Reciprocal	NDM	75.53%
LAS [29]+CCOM [28]	Texture	Brodatz	Reciprocal	NVDM	**76.43%**

recently proposed Reciprocal Rank Fusion [32] methods. In the following, we describe the rank aggregation methods and how them can be improved by our proposed estimation measures.

Both Borda [31] and Reciprocal [32] methods consider the rank information, i.e., the positions of images in ranked lists produced by different descriptors. Let $\mathcal{D}=\{D_1, D_2, \ldots, D_m\}$ be a set of CBIR descriptors and let a img_q be a query image. For each descriptor $D_j \in \mathcal{D}$, we can compute a different ranked list τ_{q,D_j} for the image query img_q. A given image img_i is ranked at different positions (defined by $\tau_{q,D_j}(i)$) according to each descriptor $D_j \in \mathcal{D}$. The objective is to use these different rank positions to compute a new distance between images img_q and img_i.

1) Borda: The Borda [31] method considers directly the rank information for computing the new distance $F_{Borda}(q,i)$ between images img_q and img_i. Specifically, the distance is scored by the number of images not ranked higher than image img_i in the different ranked lists [33]. The new distance can be computed as follows:

$$F_{Borda}(q,i) = \sum_{j=0}^{m} \tau_{q,D_j}(i). \qquad (6)$$

The Borda method does not differ between positions given by high-effective and low-effective ranked lists. In this way, we propose to use the effectiveness estimation measures for computing *weights* for ranked lists computed by each descriptor, as a linear combination. The new distance can be computed as follows:

$$F_{BordaW}(q,i) = \sum_{j=0}^{m} \tau_{q,D_j}(i) \times f_{D_j}(q), \qquad (7)$$

where the function $f_{D_j}(q)$ can be computed by the proposed NVM and NVDM estimation measures using the image representations of img_q.

2) Reciprocal Rank Fusion: The Reciprocal Rank Fusion also uses the rank information for computing a similarity score between images img_q and img_i. The scores are computed according to a naïve scoring formula:

$$F_{Reciprocal}(q,i) = \sum_{j=0}^{m} \frac{1}{k + \tau_{q,D_j}(i)}, \qquad (8)$$

where k is a constant. Analogous to Borda method, we propose to use the effectiveness estimation measures for computing weights for ranked lists computed by each descriptor. The NVM and NVDM estimation measures can be used to compute the function $f_{D_j}(q)$.

$$F_{ReciprocalW}(q,i) = \sum_{j=0}^{m} \frac{1}{k + (\tau_{q,D_j}(i) \times f_{D_j}(q))}, \qquad (9)$$

3) Rank Aggregation Results: Table II presents the results obtained for rank aggregation tasks considering the three datasets (shape, color, and texture). We present the results of original rank aggregation methods and their combination with proposed estimation measures, with best values in bold. For all datasets, the combination of proposed measures reached the best results.

V. Conclusions

In this work, we have presented two unsupervised measures for estimating the effectiveness of ranked lists on CBIR tasks. First an image representation is proposed to encode contextual information extracted from ranked lists and distance scores. In the following, visual patterns found in this image representation are used to define measures for estimating the effectiveness of ranked lists.

Experiments involving shape, color, and texture descriptors demonstrated the applicability of proposed measures. Obtained results demonstrate that our approaches can provide accurate estimations of effectiveness of ranked lists. The use of proposed measures are also evaluated on image retrieval tasks aiming at improving the effectiveness of rank aggregation approaches.

Future work focuses on: (*i*) considering the combination of different estimation measures for more accurate prediction of effectiveness; (*ii*) using proposed methods aiming at improving the effectiveness of other rank aggregation methods; (*iii*) using proposed measures on other unsupervised approaches as image re-ranking algorithms.

VI. Acknowledgments

References

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 5:1–5:60, 2008.

[2] R. da S. Torres and A. X. Falcão, "Content-Based Image Retrieval: Theory and Applications," *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.

[3] C. D. Ferreira, J. A. dos Santos, R. da S. Torres, M. A. Gonçalves, R. C. Rezende, and W. Fan, "Relevance feedback based on genetic programming for image retrieval," *Pattern Recogninion Letters*, vol. 32, no. 1, pp. 27–37, 2011.

[4] J. A. dos Santos, C. D. Ferreira, R. da S. Torres, M. A. Gonçalves, and R. A. Lamparelli, "A relevance feedback method based on genetic programming for classification of remote sensing images," *Information Sciences*, vol. 181, no. 13, pp. 2671 – 2684, 2011.

[5] P. Kontschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis," in *Asian Conference on Computer Vision*, 2009, pp. 655–666.

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

[7] J. Jiang, B. Wang, and Z. Tu, "Unsupervised metric learning by self-smoothing operator," in *IEEE International Conference on Computer Vision (ICCV'2011)*, 2011, pp. 794–801.

[8] X. Yang, L. Prasad, and L. Latecki, "Affinity learning with diffusion on tensor product graph," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 28–38, 2013.

[9] X. Yang and L. J. Latecki, "Affinity learning on a tensor product graph with applications to shape and image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011)*, 2011, pp. 2369–2376.

[10] D. C. G. Pedronette and R. da S. Torres, "Exploiting pairwise recommendation and clustering strategies for image re-ranking," *Information Sciences*, vol. 207, no. 1, pp. 19–34, 2012.

[11] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu, "Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2012)*, 2012, pp. 3013 –3020.

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

[13] O. Schwander and F. Nielsen, "Reranking with contextual dissimilarity measures from representational bregmanl k-means," in *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP'2010)*, vol. 1, 2010, pp. 118–122.

[14] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," in *Computer Analysis of Images and Patterns (CAIP'2011)*, vol. 6854, 2011, pp. 369–376.

[15] ——, "Exploiting contextual information for rank aggregation," in *International Conference on Image Processing (ICIP'2011)*, 2011, pp. 97–100.

[16] ——, "Exploiting contextual information for image re-ranking and rank aggregation," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 2, pp. 115–128, 2012.

[17] L. J. Latecki, R. Lakmper, and 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.

[18] D. C. G. Pedronette and 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.

[19] R. da S. Torres and A. X. Falcão, "Contour Salience Descriptors for Effective Image Retrieval and Analysis," *Image and Vision Computing*, vol. 25, no. 1, pp. 3–13, 2007.

[20] N. Arica and F. T. Y. Vural, "BAS: a perceptual shape descriptor based on the beam angle statistics," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1627–1639, 2003.

[21] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

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

[23] R. O. Stehling, M. A. Nascimento, and 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.

[24] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, pp. 762–768.

[25] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal on Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[26] J. van de Weijer and C. Schmid, "Coloring local feature extraction," in *European Conference on Computer Vision (ECCV'2006)*, vol. Part II, 2006, pp. 334–348.

[27] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

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

[29] B. Tao and B. W. Dickinson, "Texture recognition and image retrieval using gradient indexing," *Journal of Visual Comunication and Image Representation*, vol. 11, no. 3, pp. 327–342, 2000.

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

[31] H. P. Young, "An axiomatization of borda's rule," *Journal of Economic Theory*, vol. 9, no. 1, pp. 43–52, 1974.

[32] G. V. Cormack, C. L. A. Clarke, and 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.

[33] A. Khudyak Kozorovitsky and O. Kurland, "Cluster-based fusion of retrieved lists," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, ser. SIGIR '11, 2011, pp. 893–902.