Exploiting Contextual Information in Image Retrieval Tasks

Daniel Carlos Guimarães Pedronette, Ricardo da S. Torres
Recod Lab - Institute of Computing
University of Campinas (UNICAMP)
Campinas, SP - Brazil
{dcarlos,rtorres}@ic.unicamp.br

***Abstract*—In Content-based Image Retrieval (CBIR) systems, accurately ranking images is of great relevance, since users are interested in the returned images placed at the first positions, which usually are the most relevant ones. In general, CBIR systems consider only pairwise image analysis, that is, compute similarity measures considering only pairs of images, ignoring the rich information encoded in the relations among several images. On the other hand, the user perception usually considers the query specification and responses in a given *context*. We propose five *re-ranking* and *rank aggregation* algorithms aiming at exploit *contextual information* for improving the effectiveness of CBIR systems. Re-ranking algorithms exploit contextual information, encoded in the relationships among collection images, while rank aggregation approaches have been used to combine results produced by different image descriptors. We also propose approaches for combining the re-ranking and rank aggregation methods and for efficient re-ranking computation on GPUs.**

***Keywords*-content-based image retrieval; contextual information; re-ranking; rank aggregation; multimodal retrieval; parallel computing;**

I. Introduction

The huge growth of image collections and multimedia resources available and accessible through various technologies is remarkable. In this scenario, there is the need of methods for indexing and retrieving these data. A promising approach relies on supporting image searches by taking into account image content information, using the Content-Based Image Retrieval (CBIR) systems.

Basically, a CBIR system aims at meeting the user needs expressed in a query specification (*e.g.*, by defining a query image as input). The method usually applied to achieve its goal relies on retrieving the most similar images in a collection by taking into account image visual properties. Collection images are *ranked* in decreasing order of similarity, according to a given *image descriptor*. An image content descriptor is characterized by [1]: (*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.

Several efforts have been proposed for improving the effectiveness of CBIR approaches. Some of those efforts rely on using more accurate features for describing images. Another venue is related to the definition of similarity (or distance) functions that would be able to measure the distance between feature vectors in a more effective way.

However, in general, those iniatives perform only pairwise image analysis, that is, they compute similarity (or distance) measures considering only pairs of images, ignoring the rich information encoded in the relationships among images. On the other hand, the user perception usually considers the query specification and the query responses in a given *context*. Context can be broadly defined as all information about the whole situation relevant to an application and its set of users. In interactive applications, the use of context can play an important role. In a CBIR scenario, relationships among images, encoded in ranked lists and distances among images, can be used for extracting *contextual information* [2].

We propose five different *re-ranking* methods aiming at improving the effectiveness of image retrieval tasks replacing pairwise similarities by more global affinities. For each re-ranking method, we propose a *rank aggregation* approach that uses the re-ranking methods to combine CBIR descriptors. In summary, we put efforts on post-processing the distance/similarity scores, by taking into account the *contextual information* available in relationships among images in a given collection. The methods require no user intervention, training or labeled data, operating on an absolutely *unsupervised* way. The use of different strategies by the proposed methods contributes for analysing different, and complementary aspects encoded in relationships among images. Therefore, we also proposed approaches for combining the outputs of *re-ranking* and *rank aggregation* methods, aiming at further improving the effectiveness of CBIR results. In addition, we also proposed an efficient parallel approach for computing re-ranking methods on GPUs.

We face several relevant research challenges in order to exploit contextual information: What kind of data can be used for exploiting useful contextual information? How can we process these data in order to extract useful contextual information? How to process the minimum amount of raw data to extract the most useful contextual information? How can we combine different approaches aiming at further improving the retrieval effectiveness? How can we efficiently compute re-ranking and rank aggregation methods?

Solutions for various of these questions are addressed in this PhD work and presented along this paper. The paper is organized as follows: Section II briefly describes the problem definition. Section III presents the contributions of this PhD work. Finally, Section IV presents results and conclusions.

Due to space limitation, a detailed description of the algorithms proposed and results achieved are not presented in this paper. The full thesis [3] and related publications [4]–[16] are available in http://www.ic.unicamp.br/~dcarlos (as of 14/05/2012).

II. Image Retrieval Model and Problem Definition

Let $\mathcal{C}$=$\{img_1, img_2, \ldots, img_N\}$ be an *image collection*.

Let $\mathcal{D}$ be an *image descriptor* which can be defined [1] 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* $\hat{I}$.
- ρ: $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ *is a distance function that computes the distance between two images according to the distance between their corresponding feature vectors.*

The distance $\rho(img_i,img_j)$ among all images $img_i,img_j \in \mathcal{C}$ can be computed to obtain an $N \times N$ distance matrix A, such that $A[i,j] = \rho(img_i,img_j)$. Given an image query img_q, we can compute a ranked list R_q in response to the query, based on distance matrix A. The ranked list R_q=$(img_1, img_2, \ldots, img_N)$ can be defined as a permutation of the collection $\mathcal{C}$. We also can take every image $img_i \in \mathcal{C}$ as an image query img_q, in order to obtain a set $\mathcal{R} = \{R_1, R_2, \ldots, R_N\}$ of ranked lists for each image of collection $\mathcal{C}$.

Our goal is to propose re-ranking algorithms, represented by function f_r, that takes as input the distance matrix A (and the set of ranked lists $\mathcal{R}$ computed based on this matrix) for computing a new and more effective distance matrix $\hat{A}$:

$$\hat{A} = f_r(A, \mathcal{R}) \tag{1}$$

Given the new distance matrix $\hat{A}$, a new set $\hat{\mathcal{R}}$ can be obtained. $\hat{\mathcal{R}}$ contains the new ranking positions of all collection images, that is, the collection images are re-ranked. Note that the main aspect of f_r consists in exploiting all relationships encoded in both A and $\mathcal{R}$.

We consider an iterative approach for our re-ranking methods, in which the distance matrix A is progressively improved. An abstract iterative model of our re-ranking methods is illustrated in Figure 1.

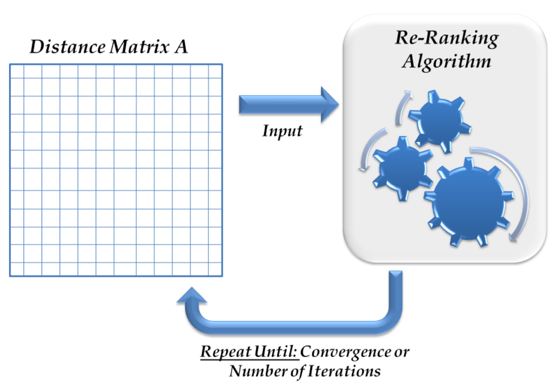

Fig. 1. Abstract model of the proposed image re-ranking algorithms.

III. Contributions

A. Distance Optimization Algorithm

The *Distance Optimization Algotihm - DOA* ([6], [8], [14]) is a re-ranking method that exploits a clustering approach for performing image re-ranking in CBIR tasks. The algorithm explores the fact that if two images are similar, their distances to other images and therefore their ranked lists should be similar as well. The main idea of the algorithm consists in clustering images and then using the created clusters for updating distances and performing image re-ranking. These steps are repeated in an iterative manner until a convergence criterion is reached.

Clusters are created according to the contextual information about images, considering the similarity of their ranked lists or the correlation of their distances. These similarity scores are computed by special functions called *cluster-similar functions*. Two approaches based on the similarity of ranked lists and distances correlations are proposed to implement cluster-similar functions.

Once clusters are created, distances among images belonging to the same cluster are updated (decreased). Redefining distances leads to performing a re-ranking of the set of ranked lists. This process is repeated until the "quality" of the formed clusters does not improve and, therefore, "good" ranked lists are created. We also demonstrate how the clusters produced by the algorithm can be applied to the problem of combining ranked lists (rank aggregation) defined by different CBIR descriptors.

B. Pairwise Recommendation

We proposed a new re-ranking method conceptually based on recommender systems. Recommender systems attempt to reduce information overload by selecting automatically items that match the personal preferences of each user. Our *Pairwise Recommendation* [4] approach is inspired by the concept of recommendation, originally created to consider users preferences. However, our method does not require any user interaction. The recommendations are simulated based on information encoded in ranked lists computed by CBIR descriptors. The relationships among images encoded in ranked lists are used for composing *image profiles* and then making recommendation.

The *Pairwise Recommendation* re-ranking algorithm is based on the conjecture that images can *recommend* images found at the first positions of their ranked lists (*e.g.*, their *k-nearest neighbors*). In this context, a recommendation means that the distance between two images should be decreased and an image should be *moved up* in the ranked list of the image that received the recommendation.

Each recommendation is associated with a different *weight* (how much the distance should be decreased). For computing the recommendation weight, we consider the position of images in ranked lists and the *quality* of the ranked lists, estimated by a proposed *cohesion* measure. We consider, in this way, that images with better ranked lists (higher cohesion) have more authority for making recommendations. After performing all recommendations, ranked lists are considered for clustering images and additional recommendations are made, given the obtained clusters.

Once all distances have been updated by recommendations, a re-ranking can be performed based on the new distance matrix. These steps are repeated in an iterative manner until a convergence criterion is reached. The employed convergence criterion is based on the variation of cohesion measure. At each iteration we also increment the number of k neighbors considered for recommendations. After each iteration, more relevant images are found at first positions of the ranked lists. Non-relevant images are moved out from the first positions of the ranked lists and therefore k can be increased for considering more images. Finally, when the convergence criterion is reached, a re-ranking is performed based on the final distance matrix computed.

Figure 2 present some results, illustrating the precision $\times$ recall curves of shape descriptors CFD [8] and IDSC [17] on MPEG-7 [18] dataset, considering different situations: before and after using the *Pairwise Recommendation* (PRM) re-ranking algorithm, and after using it for rank aggregation. As it can be observed, for both re-ranking and rank aggregation, very significant gains in terms of precision have been achieved. Our recommendation-based strategy opens a new area of investigation, related to the use of recommendation techniques in re-ranking tasks.

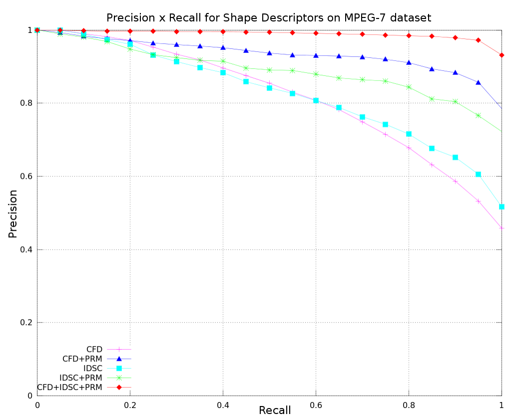

Fig. 2. Precision x Recall curves for two shape descriptors before and after the *Pairwise Recommendation* (PRM) re-ranking and rank aggregation.

C. Context Spaces

We propose the creation of *contextual spaces* [5], [10] for analysing the contextual information, *i.e.*, for characterizing the *local* context of a query image defined by its neighbourhood. A *contextual space* is a bidimensional representation of the image collection, considering two reference images.

In the proposed approach, the *contextual spaces* are constructed considering the most similar images to a given query image. Later, new distances are computed by taking into account the distances among these neighbours to other collection images. The image collection is re-ranked based on the new distances and this process is repeated along iterations.

In fact, the main motivation of the Context Spaces Re-Ranking algorithm relies on the following question: "*Given a query image, what information can similar images provide about other collection images?*" In this method, this question is answered by exploiting *contextual spaces*. The reasoning behind the use of *contextual spaces* consists in taking into account relationships of images in the context of the query (and not only pairwise distances). Given a query image, a selected set of similar images to the query is considered for constructing the *contextual spaces*. In the following, information encoded in these contextual spaces is used for computing a new distance from the query image to other collection images. In this way, distances from an image to other collection images are redefined considering the distances to their similar images. Figure 4 illustrates an example of *contextual space* for two similar reference images, given in Figure 3.

Fig. 3. Similar reference images.

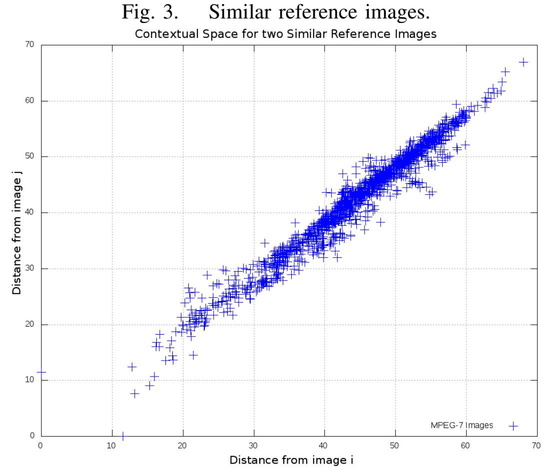

Fig. 4. Contextual space for two similar images.

Since the distances among images are redefined to become more effective, the process is repeated iteratively, aiming at further improving the effectiveness along iterations.

D. RL-Sim

We present the *RL-Sim Re-Ranking* algorithm [12], [15], a new re-ranking method that considers the similarity among ranked lists (***R****anked*__*L*__*ists*-***Sim****ilarities*) for characterizing contextual information in CBIR systems. The main motivation of this re-ranking algorithm relies on the conjecture that *contextual information encoded in the similarity between ranked lists can provide useful information for improving the effectiveness of CBIR descriptors*. In general, if two images are similar, their ranked lists should be similar as well.

We believe that the modelling of contextual information considering only the similarity between ranked lists represents

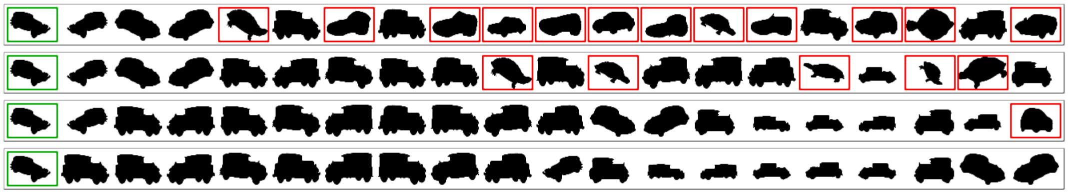

Fig. 5. Evolution of rankings along iterations using RL-Sim re-ranking algorithm on the MPEG-7 [18] dataset.

an advantage of our strategy. Since the re-ranking method does not depend on distances or similarity scores, it can be used for different CBIR tasks and can be easily adapted for other information retrieval tasks (*e.g.*, text or multimodal retrieval). Beyond that, the re-ranking method can use different similarity/distance measures among ranked lists, a well-established research area [19]. Therefore, the re-ranking algorithm can be easily extended using different similarity measures.

A contextual distance measure is defined based on similarity/dissimilarity of ranked lists. The contextual distance measure represents the basis of RL-Sim re-ranking algorithm. The distance value $\rho(img_i,img_j)$ between two images img_i, $img_j \in \mathcal{C}$ considers only the two images img_i, img_j. In order to compute the ranked lists R_i, R_j for images img_i, img_j, distances from these images to all other collection images need to be computed. In this way, the ranked lists represent, by itself, a contextual description of images with regard to the whole dataset. The images at the top positions of ranked lists often represent the most relevant images, in the sense that they usually represent the results in which users are interested. In this scenario, we conjecture that *given any two images and their respective ranked lists, a new and more effective distance measure between the two images can be computed by considering the images at the top positions of their ranked lists*.

The objective of the proposed re-ranking algorithm is to move the non-similar images down in the ranked lists, and as a result of this process, the quality of ranked lists is improved. The iterative behavior of the *RL-Sim Re-Ranking* algorithm can be observed in results illustrated in Figure 5. The figure shows the evolution of rankings (and their precision) along iterations. Each row presents 20 results for a query image (first column with green border). The first row presents the results of CFD [8] shape descriptor (wrong results with red borders). The remaining rows present the results of *RL-Sim Re-Ranking* algorithm for each iteration. We can observe that wrong results contain images from different classes, situation in which the re-ranking algorithm can *correct* the rankings based on the contextual information.

E. Contextual Re-Ranking

The *Contextual Re-Ranking* algorithm [7], [9], [11] presents a novel approach for retrieving contextual information, by creating a *gray scale image* representation of distance matrices computed by CBIR descriptors (referenced as *context image*). Our method uses distance matrices computed by CBIR descriptors that are later processed considering their image representation. The use of image processing techniques for *contextual information* representation and processing is an important novelty of our work.

The *context images* represent a great source of information about the image collection. A single *context image* contains information about all distances among images and their spatial relationship defined by the ranked lists of the reference images. An example, considering two similar reference images (from the MPEG-7 dataset [18]), is illustrated in Figure 6. The respective gray scale image representation is illustrated in Figure 7.

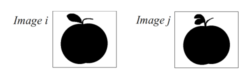

Fig. 6. Similar reference images.

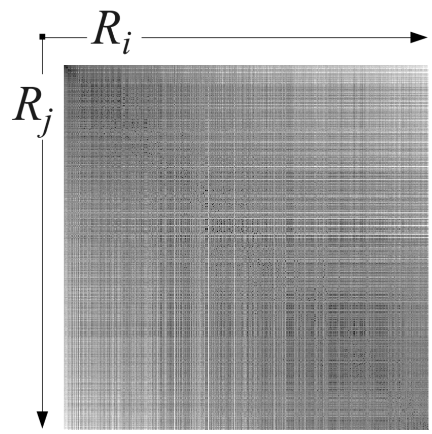

Fig. 7. Context image for similar reference images.

The *context image* is constructed for the k-Nearest Neighbors of a query image and analysed using image processing techniques. The median filter, which is a well-known non-linear filter often used for removing noise, is exploited in our approach to improve the quality of distance scores. Basically, we consider that "*wrong*" distances can be considered and represented as "*noise*" in the *context image*, and the median filter is used for filtering this noise out. In fact, a very large number of image processing techniques can be used for extracting useful information from context images. We believe that our strategy opens a new area of investigation related to the used of image processing approaches for analyzing distances computed by CBIR descriptor, in tasks such as image re-ranking, rank aggregation, and clustering.

The main idea of the algorithm consists in, given an image

$img_i \in \mathcal{C}$, to process *contextual information* of img_i by constructing *context images* for each one of its *k-nearest neighbors* (based on the distance matrix A). We use an affinity matrix W to store the results of processing *contextual information*. Let N be the size of collection $\mathcal{C}$, the affinity matrix W is an $N \times N$ matrix where $W[k, l]$ represents the similarity between images img_k and img_l.

We use image processing techniques to process the *context images* that consider img_i and each one of its *k-nearest neighbor* and then update the affinity matrix W. The same process is performed for all $img_i \in \mathcal{C}$. Since all images of $\mathcal{C}$ are processed, the affinity matrix W is used as input for computing a new distance matrix A_{t+1} (where t indicates the current iteration).

F. Efficient Image Re-Ranking Computation on GPUs

The usefulness of CBIR systems depends on both the *effectiveness* and the *efficiency* of the retrieval process. While the effectiveness is related to the quality of retrieved images, the efficiency is related to the time spent to obtain the results. Both are indispensable for useful real-world systems. Aiming at computing the relationship among images, re-ranking algorithms often consider all distances among images in a given dataset, which represent a large computational effort.

The computation of re-ranking algorithms can be parallelized using Graphic Processing Units (GPUs) devices. Despite the fact that these distances can be computed concurrently in multi-core machines and the execution is likely to achieve linear speedups, current multi-core machines are still limited to a couple of cores, which limit the maximum available parallelism. GPUs, on the contrary, are capable of executing up to 1600 threads at the same time, two orders of magnitude higher than state-of-the-art multi-core CPUs. Graphic Processing Units (GPUs) have evolved into massive parallel architectures capable of executing hundreds of operations per cycle.

In this work [13], we address the image re-ranking performance challenges by designing and implementing a re-ranking algorithm that takes advantage of the massive amount of parallelism at GPUs. We propose a parallel GPU-based solution which can speed up the *Contextual Re-Ranking* algorithm computation.

We used the OpenCL standard in our implementation. OpenCL is a new industry standard for task-parallel and data-parallel heterogeneous computing on a variety of modern CPUs, GPUs, DSPs, and other microprocessor designs [20]. In OpenCL, a program is executed on a computational *device*, which can be a CPU, GPU, or another accelerator.

The *Contextual Re-Ranking* algorithm presents great potential for parallelization, as a large number of context images ($N \times K$) are created at each iteration and they do not depend on each other for processing. The re-ranking step (the re-sort of N ranked lists) also can be computed in a parallel way for each ranked list. However, the parallelization of the algorithm presents several challenges (e.g., concurrent accesses to the affinity matrix W) that may require the design of synchronization approaches.

We compared the OpenCL implementation with the serial code in C/C++ and the original implementation in Java. We also compared the performance of the OpenCL implementation for different kernels and executing in different devices - CPU and GPUs. Experimental results demonstrate that very significant speedups (between $4\times$ and $8\times$) can be achieved by the OpenCL parallel implementation.

G. Combining Re-Ranking and Rank Aggregation

Although a lot of efforts have been employed to develop new re-ranking and rank aggregation methods, few initiatives aim at combining the existing methods. Besides that, in the same way that different CBIR descriptors produce different and complementary rankings, results of *re-ranking* and *rank aggregation* methods can also be combined to obtain more effective results.

We propose three novel approaches [16] for combining re-ranking and rank aggregation methods aiming at improving the effectiveness of CBIR systems. We discuss how to combine:

- *Re-Ranking Algorithms*: an output matrix obtained from a re-ranking algorithm can be submitted to other re-ranking algorithm as input, aiming at further improving its effectiveness. We call this combination approach as "cascading re-ranking";
- *Rank Aggregation Algorithms:* given a set of image descriptors, different rank aggregation methods can be employed for combining them. However, each rank aggregation method produces a different output. In this way, another rank aggregation method can be employed for combining the results of the first rank aggregation methods.
- *Re-Ranking and Rank Aggregation Algorithms:* a single image descriptor can be submitted to different re-ranking algorithms and a rank aggregation method can combine the results of different re-ranking algorithms in order to obtain a single, and more effective distance matrix.

IV. Results and Conclusions

In this PhD work, we have proposed different re-ranking and rank aggregation methods for exploiting contextual information in image retrieval tasks. A large experimental evaluation demonstrated that the methods can be used for general CBIR tasks considering different visual (*e.g.*, shape, color, and texture) and multimodal (visual and textual) features. Experimental results demonstrate the effectiveness of the proposed methods. We have also showed that the methods can be combined and efficiently computed using parallel computing.

This PhD work has been done in three years and has resulted in:

- Four international journal papers [4]–[7];
- Seven international conference papers [8]–[13], [16]
- One national conference paper [14];
- One paper submission to an international journal [15];

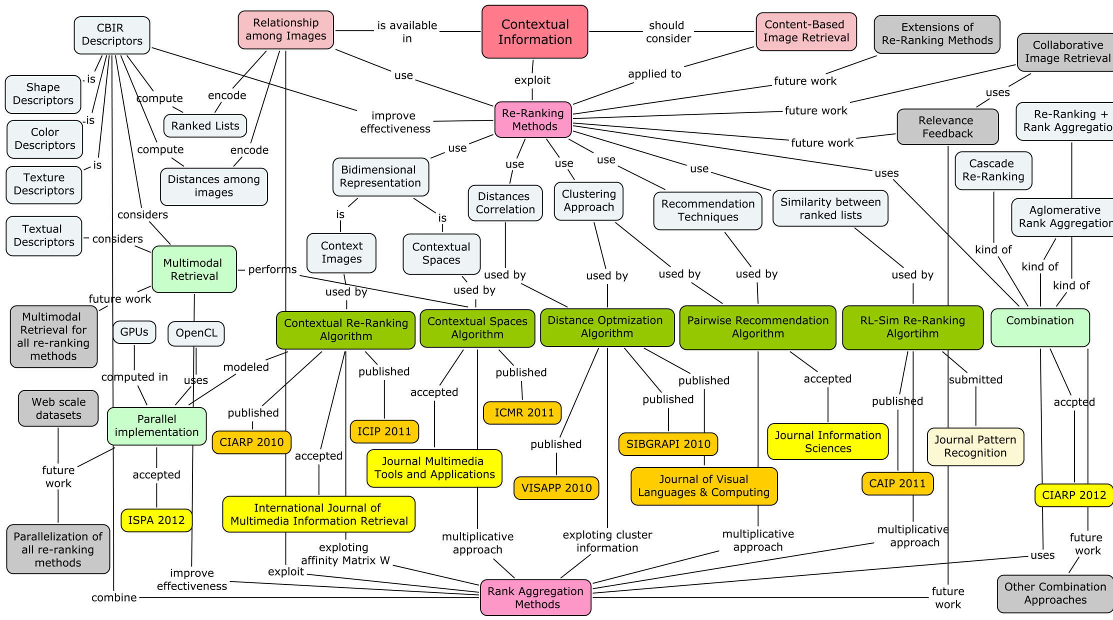

Fig. 8. Main concepts, contributions, publications, and possible extension of the PhD work.

Figure 8 summarizes the overall organization of the PhD work, considering the main concepts and their relationships. This figure also shows the main contributions, associated publications, and future work. The colors of this figure aim at organizing the meanings of each concept: in red, the main subjects related to the research; in green, the main contributions of this work; in blue, the related concepts used to address the contributions; in yellow, the associated publications; and in gray, the future work.

Acknowledgment

Authors thank AMD, FAEPEX, CAPES, FAPESP, and CNPq for financial support. Authors also thank DGA/UNICAMP for its support in this work.

References

[1] 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.

[2] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 2–11, 2010.

[3] D. C. G. Pedronette, "Exploiting contextual information for image re-ranking and rank aggregation in image retrieval tasks," Ph.D. dissertation, University of Campinas, Brazil, 2012, Advisor: Ricardo da S. Torres.

[4] D. C. G. Pedronette and R. da S. Torres, "Exploiting pairwise recommendation and clustering strategies for image re-ranking," *Information Sciences*, accepted.

[5] D. C. G. Pedronette, R. da S. Torres, and R. C. Tripodi, "Using contextual spaces for image re-ranking and rank aggregation," *Multimedia Tools and Applications*, accepted.

[6] D. C. G. Pedronette and R. da S. Torres, "Exploiting clustering approaches for image re-ranking," *Journal of Visual Languages and Computing*, vol. 22, no. 6, pp. 453–466, 2011.

[7] D. C. G. Pedronette and R. da S. Torres, "Exploiting contextual information for image re-ranking and rank aggregation," *International Journal of Multimedia Information Retrieval*, accepted.

[8] 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.

[9] D. C. G. Pedronette and R. da S. Torres, "Exploiting contextual information for image re-ranking," in *Iberoamerican Congress on Pattern Recognition (CIARP'2010)*, 2010, pp. 541–548.

[10] D. C. G. Pedronette and R. da S. Torres, "Exploiting contextual spaces for image re-ranking and rank aggregation," in *ACM International Conference on Multimedia Retrieval (ICMR'11)*, 2011, pp. 13:1–13:8.

[11] D. C. G. Pedronette and R. da S. Torres, "Exploiting contextual information for rank aggregation," in *International Conference on Image Processing (ICIP'2011)*, 2011, pp. 97–100.

[12] 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.

[13] D. C. G. Pedronette, R. da S. Torres, E. Borin, and M. Breternitz, "Efficient image re-ranking computation on GPUs," in *International Symposium on Parallel and Distributed Processing (ISPA'2012)*, 2012, accepted.

[14] D. C. G. Pedronette and R. da S. Torres, "Distances correlation for re-ranking in content-based image retrieval," in *Conference on Graphics, Patterns and Images (SIBGRAPI'2010)*, vol. 1, 2010, pp. 1–8.

[15] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," *Pattern Recognition*, submitted.

[16] D. C. G. Pedronette and R. da S. Torres, "Combining re-ranking and rank aggregation methods," in *Iberoamerican Congress on Pattern Recognition (CIARP'2012)*, 2012, accepted.

[17] 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.

[18] 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.

[19] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *ACM-SIAM Symposium on Discrete algorithms (SODA'03)*, 2003, pp. 28–36.

[20] J. E. Stone, D. Gohara, and G. Shi, "OpenCL: A parallel programming standard for heterogeneous computing systems," *Computing in Science Engineering*, vol. 12, no. 3, pp. 66 –73, 2010.