

Using Boundary Conditions for Combining Multiple Descriptors in Similarity Based Queries

Rodrigo F. Barroso¹, Marcelo Ponciano-Silva^{3,2},
Agma Juci Machado Traina², and Renato Bueno¹

¹ Computer Science Dept, UFSCar, São Carlos-SP, Brazil

² Computer Science Dept, ICMC-USP, São Carlos-SP, Brazil

³ Computer Science Dept, IFTM, Uberaba-MG, Brazil

Abstract. Queries dealing with complex data, such as images, face semantic problems that might compromise results quality. Such problems have their source on the differences found between the semantic interpretation of the data and their low level machine code representation. The descriptors utilized in such representation translate intrinsic characteristics of the data (usually color, shape and texture) into qualifying attributes. Different descriptors represent different intrinsic characteristics that can get different aspects of the data while processing a similarity comparison among them. Therefore, the use of multiple descriptors tends to improve data separation and categorization, if compared to the use of a single descriptor. Another relevant fact is that some specific intrinsic characteristics are essential for identifying a subset of the data. Based on such premises, this work proposes the use of boundary conditions to identify image subsets and then use the best descriptor combination for each of these subsets aimed at decreasing the existing “semantic gap”. Throughout the conducted experiments, the use of the proposed technique had better results when compared to individual descriptor use (employing the same boundary conditions) and to various descriptors combination without the use of boundary conditions.

Keywords: CBIR, multiple descriptor combination, similarity queries.

1 Introduction

The need for storing and manipulating non-traditional data (such as images, video and audio among others) is becoming very common in a variety of computer systems, thus managing and retrieving such data in an effective and efficient way is more necessary than ever. Such data are commonly called complex data. The similarity found between element pairs of such complex data is one of the most useful approach to manipulate such data and indicates how similar or distinct an object is in relation to another. In such similarity queries, the comparison is not done using the complex data elements directly but, instead, a set of features that were extracted from such data. Features extraction algorithms are commonly applied to data generating the so called feature vectors which are

used to represent the data in further processing. In the case of images, the most used features are related to texture, shape or color distribution. Comparison is done by applying distance functions over the feature vectors of two element pairs resulting in a numeric value that determines the distance (dissimilarity) between such elements. A basic means to compute similarity queries is employing the Content-Based Retrieval (CBR) techniques, or, in the case of images, the Content-Based Image Retrieval (CBIR). One of the major challenges found in CBIR is to reduce the semantic gap [1] which is characterized by the divergence between the low level image characteristics found by the extraction algorithms and the semantic interpretation of the same image made by a human being.

Besides finding the best combination between the features extractor and an appropriate distance function [2,3], the use of more than one intrinsic characteristic while comparing elements tend to present better results while reducing such semantic gap [4,3]. Furthermore, different subsets of data in a same larger set can be better represented by different characteristics [5,6].

Based on all principles above, this work proposes the use of boundary conditions in order to find subsets of similar data inside a larger data set thus allowing the choice of the best combination of multiple descriptors in each given subset. The expected result is a reduction on the aforementioned semantic gap while running similarity based queries against data in each subset.

The remainder of the paper is organized as follows. Section 2 presents the background and correlated work. Section 3 shows this works proposal and Section 4, in its turn, shows the experiments performed and the discussion about the results found. Finally, Section 5 present the papers final remarks.

2 Background and Related Work

2.1 Similarity Queries and Semantic Gap

Content Based Image Retrieval is supported by similarity queries which commonly make use of feature vectors. Such vectors, normally extracted in an automatic way, contain low level characteristics, such as color distribution, shape and texture, which aim at representing the image contents. To compare those images, the similarity between element pairs is calculated through the use of a distance function between each images feature vectors. A small distance value corresponds to a high similarity degree. A descriptor comprises a features extractor and a distance function [3]. The main similarity query operators [7] are kNN Queries (giving an element as query center, the k nearest neighbors are returned) and Range Queries (giving an element as query center and a radius, all elements positioned inside a given distance (radius) are returned).

There exist many difficulties inherent to content based image retrieval systems. The decision about what is found similar compared to what is found not similar can be subjective thus generating a perceptual variation among system users. The very own search by similarity mechanism using feature vectors assuming similar data would render results not compatible to reality. Such difficulties are directly tied to the semantic gap mentioned before [1].

2.2 Multiple Descriptors Combination

The multiple descriptors combination approach in CBIR has been demonstrating a great capacity for leveraging results accuracy in similarity queries [8,4,5,3] and such fact has its explanation on the complimentary nature of each images visual characteristics. By using multiple descriptors, the CBIR systems try to mimic human behavior on a similar image interpretation task where various image aspects (such as color, texture or shape) are compared simultaneously. Many works proposed different ways to perform multiple descriptors combination [4,5,9,10,8]. In [4], fractal dimension analysis is used to determine the intrinsic data correlation, then using it to normalize the contribution among the multiple descriptors on the similarity calculations between the elements.

Other proposals use a relevance feedback [9,10,8] in an attempt to capture the users similarity perception using his/her interactions in the system and reflecting them in the similarity calculation. In [8], a weighted calculation is done and each multiple descriptors weight is interactively changed according to the indication of the relevant images. In [9], together with the relevance feedback, functions are generated using a genetic programming algorithm what generates more complex functions that would better calculate each elements similarity [3]. Genetic programming algorithms are also used in [11] to combine local and global descriptors.

In its majority, the approaches for combining multiple descriptors dont consider the fact that distinct intrinsic characteristics (or their combination) can better identify different data subsets.

2.3 Boundary Conditions

The use of boundary conditions allows delimitating subsets of data inside the data as a whole. In this work, we understand boundary conditions for image similarity queries as any information associated to the images that can be used for estimating limits for subsets of images in a way that data in a subset can be better highlighted by a descriptor or a combination of specific descriptors. For medical imaging queries, for example, we can quote as boundary conditions items like the diagnose hypothesis mentioned by the medical doctor in the moment an exam is requested, radiological findings included by a radiologist, or even the exam type, among others.

In [6] a perceptual parameter was included in the medical imaging similarity queries. It has been verified that in a same database, there are image subsets that could be better represented by different individual descriptors thus increasing results precision. The authors proposed the use of triads formed by the perceptual parameter, features extractor and distance function. As boundary condition to subset limitation, the radiological finding has been used, i.e., some visual characteristic found by the specialist usually related to some diagnostic hypothesis. For each subset delimited by this boundary condition the best individual descriptor was defined experimentally. Further experimentation demonstrated that the use of this perceptual parameter had leveraged similarity queries results quality and helped reducing the semantic gap.

In [5] it had been shown that different subsets (or classes) of images were better represented by the combination of multiple descriptors with distinct weights. As illustrated in Figure 1 (adapted from [5]), queries using images from class B as query center had better precision levels when higher weight values were used to the texture descriptor than the shape descriptor. An opposite result had been found while considering only images pertaining to class A, and its similar to the average result while considering all images in the subset.

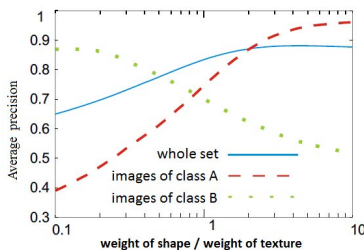


Fig. 1. Average precision with descriptors weight variation (adapted from[5])

3 Proposed Method: Multiple Descriptors Combination Using Boundary Conditions

In this work we propose the use of boundary conditions to delimitate subsets of images and, for each subset, the establishment of the best combination of multiple descriptors to be used in subsequent similarity queries.

Such descriptors combination establishment is done in an initial training phase. Later in the process, while executing a query, the boundary condition is used to identify (or estimate) the subset where the query center image is part of, so the best descriptors combination for that particular subset can be used.

A boundary condition for the larger image set must be defined first in an attempt to delimitate images subsets that have similar behavior in similarity queries. A domain application specialist is responsible for choosing such boundary condition definition considering the query objectives to be obtained while running such queries against the data set. Taking as example medical imaging applications, the boundary conditions can be defined using its associated data, like the diagnose hypothesis given by the medical specialist while requesting the exam, as well as from an initial analysis of the images provided by a specialist who can pre-classify then based in any radiological finding.

After the boundary condition is defined, the next step is to find the best descriptors combination where pre-classified images are used in this training phase and the outcome is a relation of the best descriptors combination for each boundary condition. It is worth mentioning that the criteria used as boundary condition does not need to be the same used for image classification. As a simple example, the patients sex can be used as a boundary condition while the images used for training were classified according to the illness identified in the image.

In this case, the multiple descriptors combination to be found would be the ones that better highlight the illness being considered (classes) to each image subset gathered using the boundary condition (men and women).

To demonstrate the proposed methods validity, during the training phase it has been used a simple linear combination algorithm so an exhaustive search for the best descriptors combination could be performed. Despite that, the proposed method can be used with almost any multiple descriptors combination method that can generate a valid descriptors combination in a data set.

4 Experiments

Two experiments (using two different data bases) are presented in this paper. Both image sets used were made available by the Clinical Hospital of the Medical School of Ribeirão Preto, Universidade de São Paulo, Brazil. The results evaluation was based in precision and recall curves[12].

4.1 Experiment 1: Lung Computed Tomography Exam Images

In this first experiment, a collection of 247 images related to lung computed tomography (CT) exams were classified by specialists in 6 classes according to radiologic findings: emphysema, consolidation, ground-glass opacity, interlobular septal thickening, honeycombing and normal (no findings). As boundary condition, the physicians perception has been used according to each finding: homogenous and non-homogenous increase of attenuation, inter-lobular fissures, and low attenuation with and without well defined lobule walls [6]. The extracted features were related to color [6] (high histogram, low histogram, traditional histogram) and texture (Haralick [13]).

Initially, through experimentation, descriptors with the best results were chosen (shown in Figure 2) and then the best descriptors combination for the whole set. Distinct distance function types were evaluated for each feature vector, namely the Minkowski family functions and Canberra [2]. Following, the best descriptor for each subset delimited by the boundary conditions was chosen and then triads were formed from the association between them [6]. As a final step, for each subset formed using the boundary conditions it has been calculated the best descriptors combination using the linear combination algorithm. This way, we have the final result with an ideal descriptors combination associated to each subset delimited by each of the boundary conditions.

Figure 2a) depicts the precision curves versus the recall found on the best individual descriptors for the whole set, the best individual descriptors for each boundary condition, the best combination considering the whole set and the best combination defined for each boundary condition. For the definition of the best descriptor and the best combinations, only 25% of recall was considered. The proposed technique obtained better results in all recall levels.

Figure 2b) presents the average precision for the same procedures, but now analyzing all recall levels for triad and descriptors combination definition. The

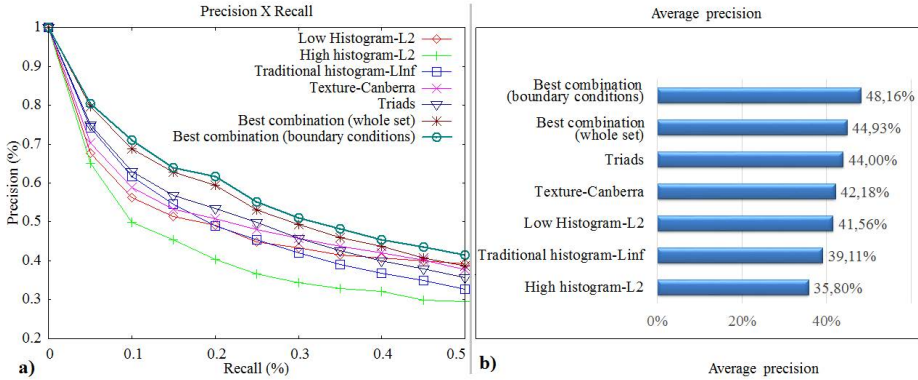


Fig. 2. CT for lung examination. a) Precision vs Recall curves. b) Average Precision. Comparison between individual descriptors, individual descriptors using boundary conditions, descriptors combination for the whole set, and descriptors combination using boundary conditions (proposed technique).

multiple descriptors combination proposed resulted in better results, with average precision next to 6.5% superior to the one obtained using descriptors combination without the boundary conditions and 14.2% better average precision obtained through the best individual descriptor use.

4.2 Experiment 2: MR Images

A set of 704 images acquired by magnetic resonance (MRI) was used to perform the second experiment, where each image was pre-classified in 40 classes according to the body region, vision plan, and cut over position they covered. As boundary condition, it has been used the exam type that generated the image: Angiogram, Axial and Coronal abdomen, Axial, Coronal and Sagittal head, Axial pelvis and Sagittal spine. The extracted features were related to color (traditional histogram), texture (Haralick) and shape (256 first Zernike moments [14]). Through the experiments, three descriptors presenting better results over the whole data (shown in Figure 3) and the triad (best descriptor associated to the boundary condition) have been defined. All recall levels in this experiment have been analyzed in order to define the triad and the best multiple descriptors combination.

Figure 3a) presents precision versus recall curves and Figure 3b) presents average precision found by each individual descriptor, individual descriptors defined for each boundary condition, best combination considering the whole set and best combination defined for each boundary condition. As it can be verified in Figure 3a), and as occurred in the first experiment, the proposed method presented better results in all recall levels and, according to Figure 3b), the average precision of the proposed method was also superior with approximately 9% gain in relation to the boundary conditions chosen for a single descriptor.

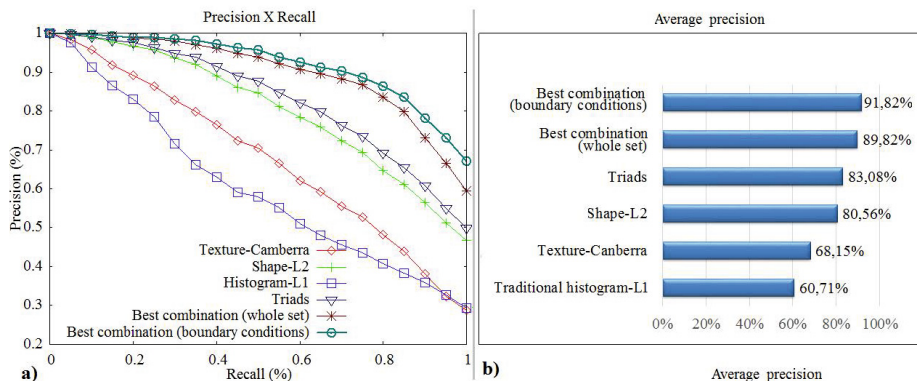


Fig. 3. Magnetic resonance images. a) Precision vs Recall curves. b) Average Precision. Comparison between individual descriptors, individual descriptors using boundary conditions, descriptors combination for the whole set, and descriptors combination using boundary conditions (proposed technique).

5 Conclusion

This paper proposed a technique that chooses the best combination of multiple descriptors for each subset in a large data set. The similar subsets were found by use of boundary conditions. This new methodology sought to simulate human perception during images analysis for purposes of comparison among them.

In the presented experiments, the proposed technique was compared with the main techniques used in CBIR found in the literature. These were: 1) use of individual descriptors for the entire set of images 2) use of a combination of descriptors for the entire set of images and 3) use of individual descriptors for each selected subsets of images, delimited by the boundary conditions. The proposed technique achieved better results in all the experiments, showing that the automatic comparison of images should consider all relevant visual aspects, each with its specific balancing to the presented context.

The use of a single descriptor for content search limited the comparison to just one criterion of similarity. On the other hand, the use of multiple descriptors for the entire set of images disregards the visual properties intrinsic to each subset delimited by the boundary conditions. Thus, the analysis presented in this paper showed a new field of research in CBIR systems, with approach to the perception of similarity of the users of these systems. The results show how promising is the use of boundary conditions for combining multiple descriptors for similarity queries on medical images.

Acknowledgments. We thank FAPESP, CAPES, CNPq and the INCT INCod for the financial support.

References

1. Deserno, T.M., Antani, S., Long, L.R.: Ontology of gaps in content-based image retrieval. *J. Digital Imaging* 22, 202–215 (2009)
2. Bugatti, P.H., Traina, A.J.M., Traina, C.J.: Assessing the best integration between distance-function and image-feature to answer similarity queries. In: SAC, pp. 1225–1230 (2008)
3. Torres, R., Falcão, A.X., Gonçalves, M.A., Papa, J.P., Zhang, B., Fan, W., Fox, E.A.: A genetic programming framework for content-based image retrieval. *Pattern Recognition* 42, 283–292 (2009)
4. Bueno, R., Kaster, D.S., Paterlini, A.A., Traina, A.J.M., Traina, C.J.: Unsupervised scaling of multi-descriptor similarity functions for medical image datasets. In: CBMS, pp. 24–31 (2009)
5. Bueno, R., Kaster, D.S., Razente, H.L., Barioni, M.C.N., Traina, A.J.M., Traina, C.J.: Using visual analysis to weight multiple signatures to discriminate complex data. In: IV, pp. 282–287 (2011)
6. Ponciano-Silva, M., Traina, A.J.M., Azevedo-Marques, P.M., Felipe, J.C., Traina, C.J.: Including the perceptual parameter to tune the retrieval ability of pulmonary cbir systems. In: CBMS, pp. 8–17 (2009)
7. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search - The Metric Space Approach. *Advances in Database Systems*, vol. 32. Kluwer (2006)
8. Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S.: Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Techn.* 8, 644–655 (1998)
9. Ferreira, C.D., dos Santos, J.A., Torres, R., Gonçalves, M.A., Rezende, R.C., Fan, W.: Relevance feedback based on genetic programming for image retrieval. *Pattern Recognition Letters* 32, 27–37 (2011)
10. Arevalillo-Herráez, M., Ferri, F.J., Domingo, J.: A naive relevance feedback model for content-based image retrieval using multiple similarity measures. *Pattern Recognition* 43, 619–629 (2010)
11. Andrade, F.S.P., Almeida, J., Pedrini, H., da S. Torres, R.: Fusion of local and global descriptors for content-based image and video retrieval. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) CIARP 2012. LNCS, vol. 7441, pp. 845–853. Springer, Heidelberg (2012)
12. Baeza-Yates, R.A., Ribeiro-Neto, B.A.: *Modern Information Retrieval*. ACM Press/Addison-Wesley (1999)
13. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics SMC-3*, 610–621 (1973)
14. Khotanzad, A., Hong, Y.H.: Invariant image recognition by zernike moments. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 489–497 (1990)