

Feature Selection Guided by Perception in Medical CBIR Systems

Pedro H. Bugatti^{1,2}, Marcela X. Ribeiro³, Agma J. M. Traina¹ and Caetano Traina Jr.¹

¹Department of Computer Science, University of Sao Paulo at Sao Carlos, Brazil

²Department of Computer Engineering, Federal Technological University of Parana, Brazil

³Department of Computer Science, Federal University of Sao Carlos, Brazil

¹{pbugatti, agma, caetano}@icmc.usp.br ³marcela@dc.ufscar.br

Abstract—This work aims at developing an efficient support to improve the precision of content-based medical image retrieval systems and also accelerate such retrieval, introducing a novel retrieval approach that integrates techniques of feature selection and relevance feedback to perform feature selection guided by perceptual similarity. Low-level features are commonly employed to represent the images by content. Feature selection is performed employing statistical association rules integrated with a relevance feedback process, tuning the mining process on the fly, according to the user's perception. This integration not only improves the feature selection accuracy, but also allows personalising such process. The experiments performed show that the method improves up to 30% the query precision and decreases up to 11.6 times the number of features employed to compute the similarity in the content-based query, also decreasing the processing costs and memory requirements of the query execution.

Keywords—Feature Selection; Relevance Feedback; User Perception; Content-Based Image Retrieval

I. INTRODUCTION

Content-based image retrieval (CBIR) techniques and similarity search have been intensively researched in the recent years. CBIR techniques work on the whole information embodied into the stored images and are not limited by a textual description of it. The advantage of working with the visual content of an image is to avoid the inherent bias and subjectivity of textual description. Textual descriptions of images can be incomplete, as the specialist can be interested in specific aspects of the image under analysis, letting some others unreported. The medical area can greatly benefit from CBIR applications due to the huge amount of images daily generated by hospitals and medical centers.

CBIR employs image processing algorithms to extract relevant characteristics (features) from the images, organizing them into feature vectors. The feature vectors are employed in place of the images to compare them, allowing fast and efficient indexing and retrieval. The features quantify intrinsic visual characteristics of the images, such as color, shape and texture [1], usually leading to high-dimensional vectors. The usage of a large number of features to represent images is a problem, because the processes of indexing, retrieving and comparing images become less accurate and more time consuming. This occurs because the usage of

a large quantity of features leads the CBIR systems to face the “dimensionality curse” problem [2]. This curse occurs because a usage of a large number of features (and consequently a high-dimensional data) leads each feature to diminish its significance. Thus, one of the major challenges in a CBIR system is to determine a minimal subset of features that best represent the images according to each specialist intention. This paper aims at dealing with this issue.

Specialists have different perceptions of image similarity, frequently feeling unfulfilled with the CBIR results. In fact, there is a disparity between low-level features and high-level user perception, which is called the *semantic gap* [3]. One way of gathering subjectivity in CBIR is to introduce the users (e.g medical specialists) into the process. To avoid the semantic gap, it is important to have the user interacting and telling what are the truly relevant images from those retrieved. This way of interaction between the user and the system is called *Relevance Feedback* (RF). The RF approach asks the specialist to quantify the relevance of each medical image retrieved by a query, allowing the system to automatically adjust future query results.

It is common that a medical image presents several clinical findings with distinct visual patterns. Thus, it is important to note that each user should perform a specific feature selection process according to his/her perception. This mainly occurs because different specialists may be interested in different clinical findings, therefore employing distinct similarity criteria. Hence, it is crucial to capture information related to the user perceptual similarity during the search process, in order to describe how each individual understands similarity in his/her specific context to perform a feature selection guided by perception.

In this paper we propose a method that integrates RF and feature selection to improve the CBIR results. In our approach, the mining process is re-executed once after the RF iterations taking into account the information fed by the specialist, reaching more precise and useful patterns.

The experiments performed show that our method improved the query precision up to 30%, yet the dimensionality reduction allowed discarding up to 91% of the features. Besides, the method speeds up the whole CBIR system,

since the complexity of the query processing decreases along the dimensionality reduction of the feature vector. Moreover, the proposed approach is effective in performing feature selection guided by each user perception.

The remainder of this paper is structured as follows. Section II summarizes the concepts and related works. Section III details the proposed method, and Section IV discusses the experiments and the achieved results. Finally, Section V presents the conclusions of the work.

II. BACKGROUND AND TECHNIQUES

Comparing images employing exact matching is not useful in several areas (e.g. medical applications), since searching for the same data already under analysis does not add extra information. In fact, the retrieval of complex data is mainly performed regarding similarity. The most well-known and useful types of similarity queries are the k -nearest neighbor and range queries.

Similarity search is performed comparing the feature vectors using a distance – or dissimilarity – function to quantify how close (or similar) is each pair of vectors. The distance functions most widely employed to perform similarity queries are those from the Minkowski family (also known as the L_p distances) [4], which are usually employed over vectorial spaces. In this paper we employed the Euclidean distance L_2 and L_1 , also called *City Block* or *Manhattan*, to perform the similarity queries.

A. Relevance Feedback

Relevance feedback fulfills an important role in CBIR systems, due to its ability to gradually reduce the semantic gap through users interaction.

As a real-time learning strategy, it can capture the essence of the user expectation, leading to higher precision and improved query refinement. Basically, a relevance feedback technique is composed of three steps in the CBIR process. Figure 1 illustrates a classic scenario of RF in conjunction with a CBIR system. In Step (1), the system retrieves the most similar images according to the initial query. In Step (2), the users guide the search process, judging the returned images based on a relevance degree; Step (3), the system captures the user's expectation based on the performed feedback and automatically adjusts the further queries based on the user's informed relevance. Steps (2) and (3) are repeated until the user is satisfied with the results. As the system captures the user's intention when a new query is performed, the resulting set of images can be continually improved until be stagnated [5].

In this work, we use the query point movement (QPM) [6] technique to perform RF, which consists of estimating the query point according to the user's interaction by moving it towards the positive examples (i.e. objects selected as relevant) and away from the negative examples (i.e. objects

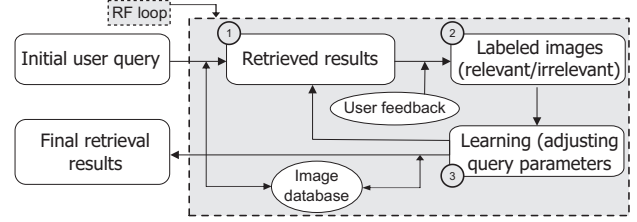


Figure 1. RF in conjunction with CBIR.

selected as irrelevant). Thus, at each user interaction cycle, an ideal query center is estimated in the query space.

The most well-known and widely used technique to iteratively improve the query center estimation through relevance feedback is the Rocchio's formula [7]. This formula is formally defined in Equation 1, where Q is the original query (i.e. the initial feature vector), Q' is the updated query (i.e. the resultant feature vector), D'_R and D'_N are respectively the positive and negative examples given by the user, $N_{R'}$ and $N_{N'}$ are the number of relevant (i.e. positive) and irrelevant (i.e. negative) examples in D'_R and D'_N , and finally α , β and γ are weighting factors, experimentally obtained, to tune the relevance feedback factors. In our approach, we employ the Rocchio's technique together with incrementally feature selection to maximize the precision of content-based image retrieval.

$$Q' = \alpha Q + \beta \left(\frac{1}{N_{R'}} \sum_{i \in D'_R} D_i \right) - \gamma \left(\frac{1}{N_{N'}} \sum_{i \in D'_N} D_i \right) \quad (1)$$

B. Statistical Association Rule Mining

In this work, we employed the StARMiner algorithm [8], which extended statistical rule mining techniques presented in [9] to find patterns in image datasets.

Let T be an image dataset, x an image class, $T_x \in T$ the subset of images of class x and $f \in F$ a feature, where F is the set of all original features employed to represent the images. Let $\mu_f(Z)$ and $\sigma_f(Z)$ be, respectively, the mean and standard deviation of the values of the feature f in the subset of images Z . StARMiner uses three thresholds defined by the user: $\Delta\mu_{min}$ – the minimum allowed difference between the average of the feature f in images from class x and the average of f in the remaining dataset; $\Delta\sigma_{max}$ – the maximum standard deviation of f values allowed in a given class and; γ_{min} – the minimum confidence to reject the hypothesis H_0 . The StARMiner algorithm mines rules of the form $x \rightarrow f$, given the following conditions are satisfied:

$$\mu_f(T_x) - \mu_f(T - T_x) \geq \Delta\mu_{min} \quad (2)$$

$$\sigma_f(T_x) \leq \Delta\sigma_{max} \quad (3)$$

$$H_0 : \mu_f(T_x) = \mu_f(T - T_x) \quad (4)$$

Equation 4 states that H_0 should be rejected with a confidence equal to or greater than γ_{min} , in favor of the hypothesis that the difference between the means $\mu_f(T_x)$ and $\mu_f(T - T_x)$ are statistically significant. A rule $x \rightarrow f$ returned by the algorithm relates a feature f with a class x , where values of f points to a statistically significant difference in the images of class x . This property indicates that feature f can distinguish images of class x . A set of rules $x \rightarrow f$ for the same predecessor x can be summarized in the rule $x \rightarrow R$, where R is the union of the features returned by StARMiner for the predecessor x , meaning that the features in the set R are the ones that can best distinguish images of class x from the remaining images. Hence the relevant features for class x are the features R . The union set S of all sets R returned by the StARMiner algorithm for all image classes contains the features that best discriminate the images into categories. The set S is the resulting set of relevant features selected by the StARMiner algorithm.

In our approach, we employ StARMiner to select the set of features that is considered in the similarity computation during the query execution. At each relevance feedback iteration, the mining algorithm is re-executed considering the feedback information fed by the user, increasing its precision and improving the feature selection result. The complexity of StARMiner is $\Theta(ckN)$, where N is the number of instances of the dataset, k is the number of features, and c is the number of categories. This algorithm was chosen because it produced better results when compared with other traditional feature selection algorithms, such as the Relief [10] and the Decision Tree Method (DTM) [11].

III. FEATURE SELECTION GUIDED BY PERCEPTION

We propose a method to improve the user satisfaction with the CBIR system that combines relevance feedback and feature selection. The relevance feedback process iteratively enhances the training dataset applied to the feature selection process according to the user intention, providing for each one a personalized training dataset closer to his/her perception. The mining process is performed once as the dataset was tuned by the user expectations.

Consider G an image database, composed of images from different classes ($n_{classes}$). Every image $f \in F$ corresponds to a feature vector. Therefore, the distance between two images is defined as $d(q, g)$ where q is a query center. Given an initial k-NN query, the retrieved images are the ones closest to the query center. However, these results may not reach the user's expectation, due to the semantic gap problem. Thus, to include the users' expectation in the process we employed the QPM relevance feedback technique, where the user will label the relevant and irrelevant images from the retrieved ones (*result* - an ordered list of the k most relevant images in G). Algorithm 1 describes the pipeline of the proposed method, while the Algorithm 2 describes the

process of incrementally building the training set according to the user's expectation.

Algorithm 1 Pipeline of the proposed method.

Require: query center q , number of k -nearest neighbors, a feature extraction function v , a distance function d , an image database G , a desirable number $N_{R'}$ relevant images to be selected, a desirable number $N_{N'_i}$ irrelevant images to be selected

Ensure: a list of *selectedFeatures*

- 1: initialQuerykNN(in:center- q , out:*result*)
- 2: RelevanceFeedback(in|out:*result*, out: $N_{N_{distinct}}$)
- 3: BuildTrainingSet(in:*result*, in: $N_{R'}$, in: $N_{N'_i}$, in: $N_{N_{distinct}}$, in: $n_{classes}$, out:*trainingSet*);
- 4: MiningRules(in:*trainingSet*, out:*selectedFeatures*)

Algorithm 2 BuildTrainingSet – Performs the incremental construction of the training set according to user feedback.

Require: a desirable number $N_{R'}$ relevant images to be selected in a given relevance cycle, a desirable number $N_{N'_i}$ irrelevant images to be selected in i -th class, a number $N_{N_{distinct}}$ of irrelevant images selected from different classes

Ensure: *trainingSet*

- 1: **if** ($N_{N'_i} < N_{R'}$) **then**
- 2: *trainingSet* += ($N_{R'} - N_{N'_i}$) irrelevant images from i -th class in result
- 3: **else if** ($N_{R'} - N_{N'_i} \notin result$) **then**
- 4: *trainingSet* += randomSelection(G , $N_{R'} - N_{N'_i}$, i)
- 5: **end if**
- 6: **if** ($n_{classes} > N_{N_{distinct}}$) **then**
- 7: **if** (*classesNotSelected* $\in result$) **then**
- 8: *trainingSet* += $N_{R'}$ relevant images
- 9: **else**
- 10: *trainingSet* += randomSelection(G , $N_{R'}$, *classesNotSelected*)
- 11: **end if**
- 12: **end if**

The steps of the proposed method are illustrated in Figure 2. The proposed method is composed of five main steps: (1) initial query, (2) relevance feedback loop (RF loop), (3) relevance feedback mining data, (4) association rule mining, (5) feature selection, which removes noisy features and bursts the most relevant ones. When the user decides to finalize and visualize the answer, similarity queries are executed with the selected features.

Initially the training dataset used in the process of association rule mining is empty. At Step 1 the user performs an initial similarity query using a given image as query center. In Step 2 the user labels the retrieved images as relevant or irrelevant. Then in Step 3 the labeled images are used by Algorithm 2 to build a training dataset.

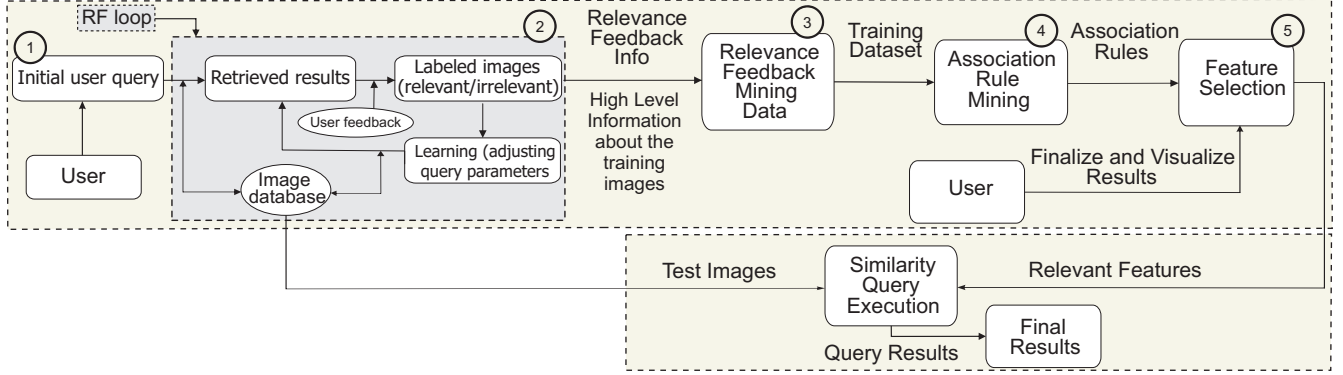


Figure 2. Pipeline of the proposed method

In Step 4, the association rule mining algorithm takes the built training dataset as input and then mine rules meeting the required support and confidence. Finally, in Step 5, the association rules generated are used in the feature selection process in order to select the most relevant features. The feature selection is accomplished according to the user expectation, since the training dataset is iteratively built during the relevance feedback performed in Step 2.

Since the user remains in the feedback loop of Step 2 until he/she is satisfied with the retrieved results, the training dataset applied to the process of association rules mining continues to be refined until the relevance feedback process reaches a saturation point. Thus, each new feedback cycle can refine the training dataset, and consequently also the feature selection process.

In order to compose the training dataset from the relevance feedback iteration, we adopted a particular strategy, which was formally described by Algorithm 2. At each relevance cycle the user is asked to select a fixed number of relevant images ($N_{R'}$), and a fixed number of irrelevant images ($N_{N'}$).

For example, in the first relevance cycle (*Cycle1*), the user could be asked to select the 5 most relevant images (i.e. those within the same class of the query image - $N_{R'} = 5$) and 1 irrelevant image (i.e. in a class distinct from the query image - $N_{N'_i} = 1$, where i is the i -th image class from the dataset).

Since we construct a balanced training dataset, when the user makes the choice and decides to feedback the system with them, the irrelevant image is used to choose the other irrelevant images (e.g. 4 images) of its same class (i -th image class) from the resultant set (*result*) of k images of the initial query to compose the training dataset too. For instance, considering an Image dataset G , if the selected irrelevant image belongs to the i -th class, the other 4 images will be from this same class.

However, there are some situations to consider. Suppose there are insufficient images from the same class (e.g class A) of the irrelevant ones in the resultant set (*result*) to complete the required number of images (i.e. $N_{R'} - N_{N'_i}$),

and to build a balanced training dataset. Then, the remaining images that belong to the same class (e.g class A) of the irrelevant one are randomly selected from the image dataset. Another situation occurs when the number of dataset classes ($n_{classes}$) is greater than the number of irrelevant images selected from distinct classes ($N_{N'_{Distinct}}$). In this case, the image classes that were not selected by the user will be randomly selected from the k resultant set (*result*) until reaching the required number (i.e. $N_{R'} - N_{N'_i}$). If it is not achieved, the remaining images are randomly selected from the image dataset. For example, considering that only one image from the i -th class was selected as irrelevant, the other irrelevant images from different classes are selected according to the described process.

It is important to emphasize that repeated images (i.e. those previously selected in past relevance cycles) are never included in the training dataset, in order to not build a biased training dataset. This fact may occur when a user in different feedback cycles selects the same image of past cycles as relevant/irrelevant.

Our method includes the high-level knowledge provided by the user from the relevance feedback process into the training dataset used by the association rule mining algorithm, thus helping to reduce the semantic gap and optimizing the feature selection process. Moreover, the method allows the system personalization, since some specialists may have different expectations regarding an image domain, thus building different training datasets.

Another point is that our method not only identifies the most relevant features according to the specialist expectation, but also performs dimensionality reduction of image features avoiding the “dimensionality curse” problem. The features are modified during each relevance feedback cycle by selecting the most significant ones and discarding the irrelevant ones. To do so, the training dataset employed to the mining of statistical association rules is incrementally increased according to the user expectation captured in each feedback cycle.

It is important to highlight that although in this paper we employ the StARMiner algorithm (see Section II-B) to

perform the association rules mining process, without loss of generality other mining methods of association rules can be applied.

IV. EXPERIMENTS

This section presents experimental results when using our proposed method to execute similarity queries. In order to evaluate our approach, we employed precision \times recall ($P \times R$) graphs [12]. The precision of a query is the fraction of the retrieved elements which are relevant. Hence, $precision = \frac{|R_A|}{|A|}$, where $|R_A|$ is the number of relevant images retrieved and $|A|$ is the size of the answer set. The recall of a query is the fraction of the relevant elements which has been retrieved. Thus, $recall = \frac{|R_A|}{|R|}$, where $|R_A|$ is the number of relevant images retrieved and $|R|$ is the number of relevant elements that should be retrieved. When analyzing $P \times R$ graphs, the closer the curve to the top, the better the corresponding retrieval technique is.

To build the $P \times R$ graphs, when the user decides to finalize and visualize the results we applied sets of k -nearest neighbor (kNN) queries, using all images of the analyzed image dataset as query centers, varying the values of k , in order to compare our approach (i.e. features selected according the user perception) with the original features.

In the Step 2 (RF Loop) of the proposed method was employed the Rocchio's technique, setting the positive feedback constant as $\beta = 1.0$ and the negative feedback constant as $\gamma = 0.5$. We chose to employ this well-known technique, for showing our method applicability. However, other more sophisticated RF methods could also be used.

A. Dataset Description

We used a variety of image datasets acquired from our Hospital University to perform the experiments. However, due to space limitations, we present here only the results obtained from one representative dataset.

The *MRI* dataset, consists of 800 images obtained from magnetic resonance imaging exams. The dataset was divided into 4 classes, according to the body region, image orientation and the height of the slice. Figure 3 illustrates an example of each class: (a) Axial Head MRI (200 images), (b) Coronal Head MRI (200 images), (c) Sagittal Head MRI (200 images), (d) Sagittal Spine MRI (200 images).

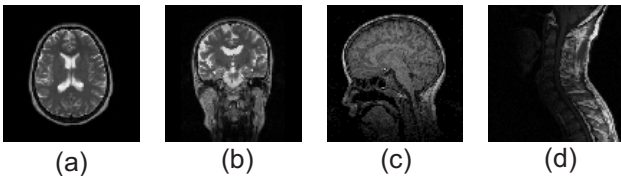


Figure 3. Examples of images from the *MRI* dataset.

Each image from the dataset was processed using the Haralick's Texture [13], the Traditional Gray-level histograms

and the Zernike Moments [14]. These feature extractors generated feature vectors comprising 140, 256 and 256 positions respectively. They were used because they describe the gray-level distribution, the texture patterns and shape peculiarities of the medical images [15].

B. Results

In our experiments, considering the Steps 1 (initial query) and 2 (RF Loop) we applied sets of k -nearest neighbor (kNN) queries, more specifically 80-nearest neighbor queries and the first three relevance feedback cycles in order to build the training dataset.

For our experiments, at the first, second and third relevance feedback cycles, 5, 10 and 15 relevant images were selected, respectively. For comparison purposes, we always required the user to perform 3 relevance cycles.

Therefore, the first relevance feedback cycle always generated a training set with $5 \times n_{classes}$, where $n_{classes}$ is the number of classes in the dataset. The second cycle generated a training set with $15 \times n_{classes}$, because it was increased by 10 new images, and thus the third RF cycle always generated a training set with $30 \times n_{classes}$ images, as it was increased by 15 new images. In all experiments, a weight $w = -0.5$ was assigned to the irrelevant images.

The $P \times R$ graphs of Figures 4, 5 and 6 show the experiments performed over the *MRI* dataset represented by the Haralick features, the gray-level histogram features, and by the Zernike moments, respectively, using L_2 distance. The graphs compare the original features (i.e. 140 features for Haralick and 256 for gray-level histogram and Zernike moments) with the proposed method refining the training dataset using the first 3 relevance cycles, gradually decreasing the number of features by selecting the relevant ones according to the user perception, and at the same time obtaining a gain in the precision of the similarity queries.

Figure 4 shows the $P \times R$ graphs obtained using the Haralick features over the *MRI* image dataset. Analyzing the graphs of Figure 4, we observe that the proposed method not only clearly improves the precision of the similarity queries, but also achieved a considerable dimensionality reduction. The precisions obtained by the relevant features selected using the training set generated in cycles 1, 2 and 3 show that the curves practically ties, presenting a precision gain of up to 30% over the precision of the original features (i.e. 140 features) for a recall level of 40%. It is important to note that the proposed method not only presents a considerable precision gain in all cases but also promoted a significant reduction of up to 6.7 times less dimensions in the third interaction, which reduces the memory and processing cost. These results show that the proposed technique improves the precision of similarity queries, even when it reduces the dimensionality of the feature vectors performing the feature selection guided by the user perception.

Figure 5 shows the P×R graphs obtained from the gray-level histogram features of the *MRI* dataset. Analyzing Figure 5, we see that our method again presented a gain in precision of up to 14% for a recall level of 20% in comparison with the original features and those selected by the mining process using the third refined training set. In this case the feature vector size were approximately reduced to up to 11.6 times less dimensions in comparison with the dimension of the complete feature vectors.

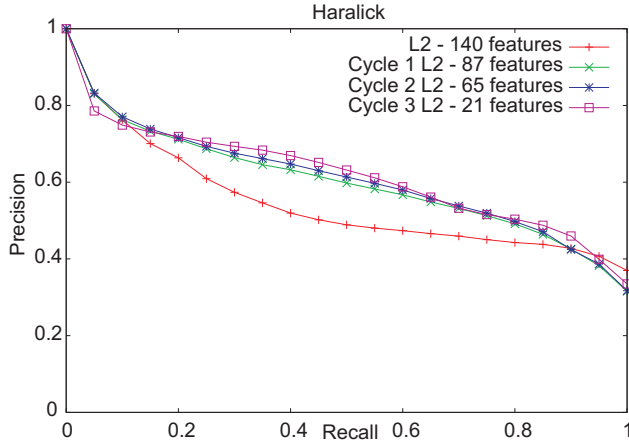


Figure 4. P×R graphs using L_2 distance obtained over the *MRI* dataset represented by Haralick features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

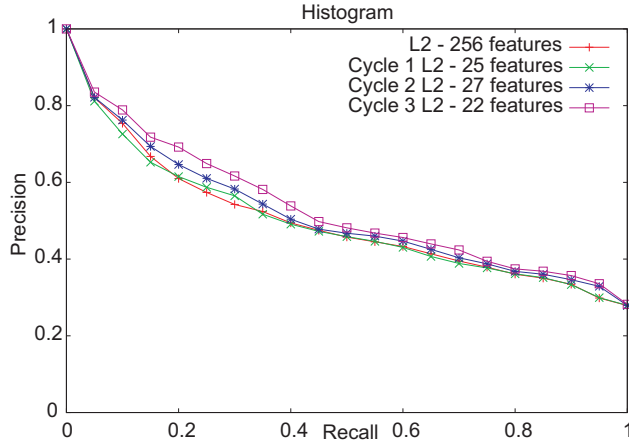


Figure 5. P×R graphs using L_2 distance obtained over the *MRI* dataset represented by Gray-level Histogram features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

Figure 6 illustrates the results over the Zernike moments features. We can notice that the proposed method achieved a precision gain of up to 11% in comparison with the original features, considering a recall level of 60%. It is important to highlight that the feature vector size again was reduced to up to 2.2 times less dimensions in comparison with original ones (i.e. 256 features).

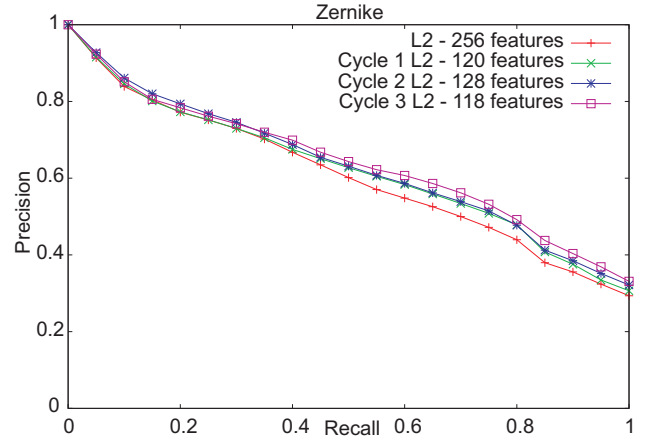


Figure 6. P×R graphs using L_2 distance obtained over the *MRI* dataset represented by Zernike Moments features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

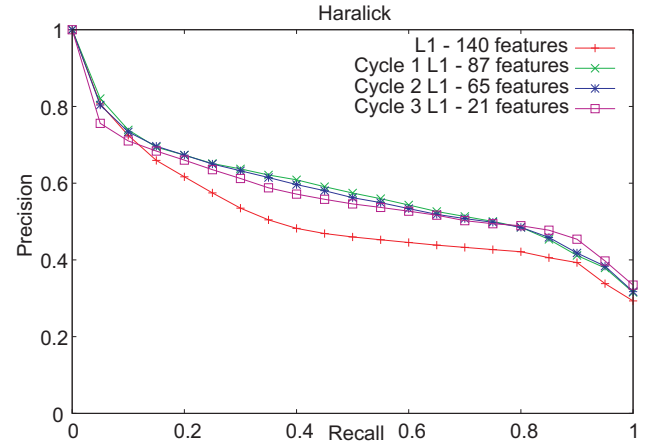


Figure 7. P×R graphs using L_1 distance obtained over the *MRI* dataset represented by Haralick features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

Figures 7, 8 and 9 show the P×R graphs obtained from the Haralick features, gray-level histogram features, and Zernike moments of the *MRI* dataset, using L_1 distance. Analyzing the graph of Figure 7, we observe that the proposed method not only clearly improves the precision of the similarity queries, but also achieved a considerable dimensionality reduction. The precisions obtained by the relevant features selected in cycles 1, 2 and 3 show that the curves practically ties, presenting a gain of 20% over the precision of the original features (i.e. 140 features) for a recall level of 40%.

Analyzing Figure 8, one can see that our method again present a gain in precision of up to 9% for a recall level of 25% in comparison with the original features, considering the third interaction. Figure 9 illustrates the results over the Zernike moments features. We can notice that in Figure 9 the method achieved a precision gain of up to 11% in comparison with the original features, considering a recall level of 50%.

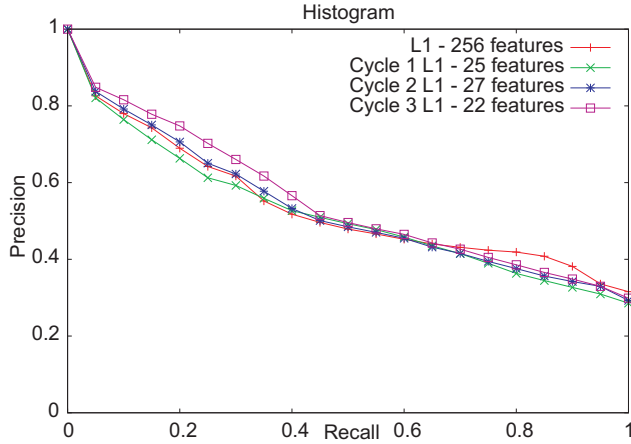


Figure 8. P×R graphs using L_1 distance obtained over the *MRI* dataset represented by Gray-level Histogram features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

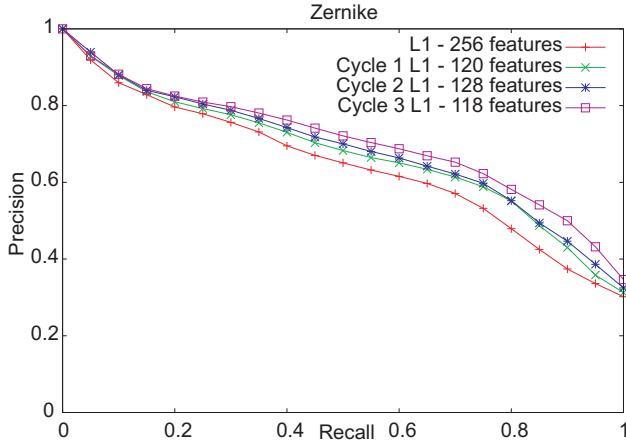


Figure 9. P×R graphs using L_1 distance obtained over the *MRI* dataset represented by Zernike Moments features comparing the retrieval ability both of the original features and of our proposed method after 3 cycles of relevance feedback.

Figure 10 shows an example of a k -NN ($k=7$) query execution, using the top left image as the query center. Figure 10(a) shows the result using the original features, and Figure 10(b) shows the result using our proposed method with fewer features. The images highlighted by a dashed line means false positive images. A false positive image is a returned image whose class differs from the class of the query center. Clearly an improvement on the results was reached when applying our method.

We also performed analyses based on the time requirements to execute similarity queries since our method reduces in a great extent the dimensionality of the features vectors. All values presented were obtained by the average of 680 queries over the *MRI* dataset with $k=200$. We performed several tests under Windows Vista 32bit, running on a machine equipped with an Intel Core 2 Duo 2.00GHz processor, 3GB of RAM and a SATA HD of 320GB and 5.400RPM. Figure 11 presents the results of these experiments consid-

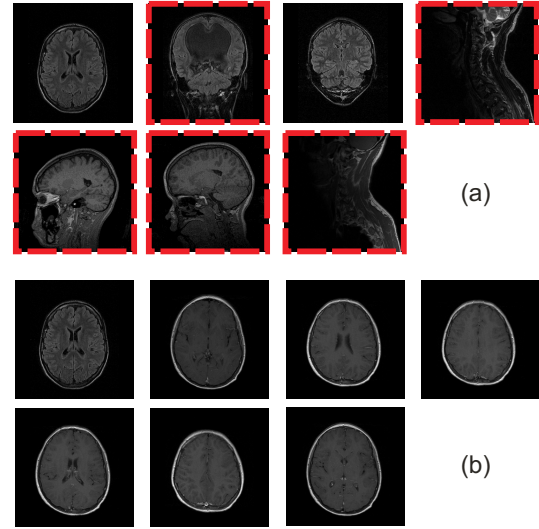


Figure 10. An example of k -NN ($k=7$) query using the top left image as query center. (a) Using the original features. (b) Using our proposed method. The images wrapped by a dashed line means false positive images.

ring: (a) Haralick features (texture), (b) histogram features (color), and (c) Zernike features (shape).

Analyzing the curves of Figure 11 (a) we can notice that the proposed method performed the queries about 6 times faster, considering the third feedback iteration in comparison with the time obtained by the original features. Figure 11 (b) illustrates a higher gain in the queries time execution, up to 10 times faster when compared with the time required by the queries using the original features. Finally, in Figure 11 (c) the proposed method again reached a considerable gain in time execution, up to 4.1 times faster in comparison with the time required by the original features.

Therefore, notable gains were also accomplished regarding not only in terms of the precision of the similarity queries, but also with respect of the time execution of them.

V. CONCLUSION

In this paper we proposed a new approach to improve the CBIR quality, dealing with the “dimensionality curse” and the “semantic gap problems”. It gathers from the users their intentions during the similarity queries and performs a feature selection guided by user perception. By coupling the relevance feedback and the mining processes, we maximized the accuracy of the feature selection process, gathering the user in the CBIR process performing feature selection guided by the user’s perception.

The feature selection performed allowed the reduction of the feature vector up to 11.6 times less dimensions (from 256 to 22), tuning the mining process according to the specialist expectation. Moreover, for each user a specific training set was built according to his/her expectation providing the personalization of the system. A simple and well-known RF method was employed, only to show the proposed method

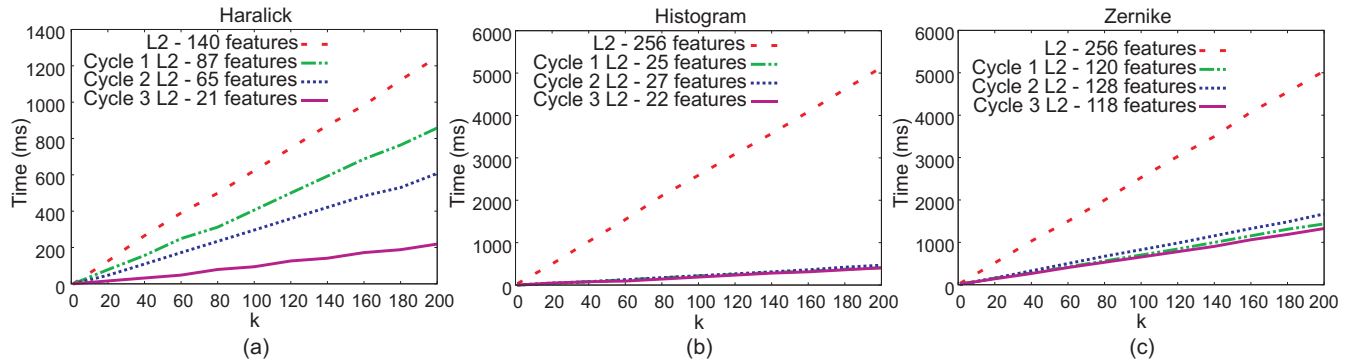


Figure 11. Graphs of k -NN queries execution times using L_2 obtained over the *MRI* dataset represented by (a) Haralick features; (b) histogram features; and (c) Zernike features, comparing the time execution between the original features and employing our proposed method applying 3 cycles of relevance feedback.

power. The use of more robust RF methods could increase the gain further. Analyses done also show that the number of three RF cycles is the most common, since after that a saturation point is reached, and the gain practically remains the same.

The experiments showed that the proposed method is effective in improving the query precision, contributing to bridge the semantic gap and achieving improvement in the query results of up to 30%. Moreover, the proposed approach can be straightforwardly extended to other types of relevance feedback techniques, feature selection methods and distance functions.

ACKNOWLEDGMENTS

This work has been supported by FAPESP (Sao Paulo State Research Foundation), CNPq (National Council for Scientific and Technological Development), CAPES (Brazilian Federal Funding Agency for Graduate Education Improvement) and Microsoft-Research.

REFERENCES

- [1] S. Jeong, S.-W. Kim, and B.-U. Choi, "Dimensionality reduction in high-dimensional space for multimedia information retrieval," in *Proceedings of the International Conference on Database and Expert Systems Applications*. Regensburg, Germany: Springer Berlin / Heidelberg, 2007, pp. 404–413.
- [2] K. S. Beyer, J. Godstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *Proceedings of the International Conference on Database Theory*, vol. 1540. Jerusalem, Israel: Springer Verlag, 1999, pp. 217–235.
- [3] T. M. Deserno, S. Antani, and R. Long, "Ontology of gaps in content-based image retrieval," *Journal of Digital Imaging*, vol. 1, no. 1, pp. 1–14, 2008.
- [4] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, no. 1, pp. 1–34, 1997.
- [5] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition Letters*, vol. 40, no. 1, pp. 262–282, 2007.
- [6] N. Doulamis and A. Doulamis, "Evaluation of relevance feedback schemes in content-based in retrieval systems," *Signal Processing: Image Communication*, vol. 21, no. 4, pp. 334–357, 2006.
- [7] J. J. Rocchio, *Relevance Feedback in Information Retrieval*, ser. The SMART Retrieval System: Experiments in Automatic Document Processing. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.
- [8] M. X. Ribeiro, A. G. R. Balan, J. C. Felipe, A. J. M. Traina, and C. Traina Jr., "Mining statistical association rules to select the most relevant medical image features," in *Proceedings of the International Workshop on Mining Complex Data*. Houston, USA: IEEE Computer Society, 2005, pp. 91–98.
- [9] Y. Aumann and Y. Lindell, "A statistical theory for quantitative association rules," *Journal of Intelligent Information Systems*, vol. 20, no. 3, pp. 255–283, 2003.
- [10] I. Kononenko, "Estimating attributes : Analysis and extension of relief," in *Proceedings of the European Conference on Machine Learning*. Catania, Italy: Springer Berlin / Heidelberg, 1994, pp. 171–182.
- [11] C. Cardie, "Using decision trees to improve case-based learning," in *Proceedings of the International Conference on Machine Learning*. Amherst, USA: Morgan Kaufmann, 1993, pp. 25–32.
- [12] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- [13] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [14] A. Khotanadz and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.
- [15] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.