On the Effects of Distance Functions to Improve Content-based Image Retrieval*

Pedro H. Bugatti[1], **Agma J. M. Traina (advisor)**[1]

[1]Department of Computer Science (ICMC)
University of São Paulo at São Carlos (USP)
Av. do Trabalhador Sãocarlense, 400, 13560-970 – São Carlos, SP – Brazil

{pbugatti,agma}@icmc.usp.br

Abstract. *The retrieval of images by visual content relies on a feature extractor to provide the most meaningful intrinsic characteristics (features) from the data, and a distance function to quantify the similarity between them. A challenge in this field supporting content-based image retrieval (CBIR) to answer similarity queries is how to best integrate these two key aspects. There are plenty of researching on algorithms for feature extraction of images. However, little attention have been paid to the importance of the use of a well-suited distance function associated to a feature extractor. This Master Dissertation was conceived to fill in this gap. It was also proposed a new technique to perform feature selection over the feature vectors, in order to improve the precision when answering similarity queries. This work also showed that the proper use of a distance function effectively improves the similarity query results. Therefore, it opens new ways to enhance the acceptance of CBIR systems.*

1. Introduction

The retrieval of multimedia data relies on a feature extractor to provide the intrinsic characteristics (features) from the data, and a measure to quantify the similarity between them. A challenge in multimedia database systems is how to best integrate these two key aspects in order to improve the quantity of the retrieved selection when answering similarity queries.

The volume of multimedia and complex data (images, videos, audio, time series, DNA sequences, among others) generated or managed in the nowadays computational systems grows in a very fast pace. It is drastically increasing not only in the amount of data, but also in the number and complexity of attributes.

Focusing this fast growing on large-scale image repositories, we clearly notice that manual annotation of images has become unfeasible, due to its inherent drawback of subjectivity, as well as non-scalability and non-uniformity of vocabulary. Content-based image retrieval (CBIR) systems are proposed to overcome these limitations, where the most similar images of a given one are retrieved based on comparisons of visual features (automatically extracted from images).

For almost a decade, researchers have been exploiting the CBIR area. On the other hand, the majority of these researches focus on the proposal of new feature extractors and

*This work has been supported by FAPESP. The Master's Dissertation is available for download at: http://www.teses.usp.br/teses/disponiveis/55/55134/tde-06052008-095020/

neglects the relationship between the features extracted from the data and the distance function employed to compare them. Hence, they make use of the most well-known and widely used distance functions, such as the Euclidean distance, which usually does not deliver a desirable similarity assessment.

In this paper we show that a careful choice of a distance function considerably improves the retrieval of multimedia data, and, to do so we analyze and compare the association and dependencies between the distance functions and the features extracted. Our claim is that precise results would only be obtained if you have the proper features extracted from the image being analyzed and a well-suited distance function to compare them.

It is important to highlight that despite the recent research efforts, CBIR remains a challenging task, due largely to the so-called "semantic gap" problem, where the low-level features automatically extracted from images do not satisfactorily represent the semantic interpretation of the images in terms of the user perception [Deserno et al. 2007]. Thus, besides the focal point of the Master's Dissertation summarized here, we not only proposed new techniques that deal with the semantic gap problem, but also performed a dimensionality reduction on the feature vectors. This is necessary because CBIR uses intrinsic visual features of images, such as color, shape and texture yielding vectors with hundreds or even thousands of features, leading to the so-called "dimensionality curse" problem [Malcok et al. 2006].

The remainder of this paper is structured as follows. Section 2 summarizes the contributions achieved from the MSc. research. Section 3 presents the conclusions, while Section 4 discusses some future directions.

2. Key Contributions

In this section, we present the key contributions achieved in the order they appeared in the original dissertation document. Due to space limitations, in the present paper we do not provide extended details about each contribution. Please check the dissertation document [Bugatti and Traina 2008] for the details about the algorithms and implementations related to each contribution.

2.1. Assessing the Best Integration between Distance-Functions and Features

There is a close relationship between the features and the distance function used to compare the data, in order to return what the human beings would expect from such comparison. However, the majority of the works concerning indexing and retrieval of multimedia data overlook this relationship and go for the most known and used distance functions, such as the Euclidean or other members of the L_p family, relegating the distance function to a secondary importance. It is important to highlight that the efficiency and the efficacy of an image retrieval technique is significantly affected by the inherent ability of the distance function to separate data. Considering two feature vectors $F = \{f_1,..., f_n\}$ and $G = \{g_1,..., g_n\}$, some important distance functions identified by this work are summarized in Table 1.

The feature vectors extracted from the multimedia data is the other key aspect for the similarity comparison between complex data. We employed a variety of features extractors based on color, texture and shape from several datasets to perform the

Table 1. Descriptions of some relevant distance functions

Name	Equation	Usage
Minkowski Family L_p	$L_p(F,G) = \sqrt[p]{\sum_{i=1}^{n} \lvert f_i - g_i \rvert^p}$	The members of the L_p distance are widely employed in the literature. The *Euclidean* distance (L_2) corresponds to the human-being notion of spatial distance. The L_1 distance (*City Block* or *Manhattan*) corresponds to the sum of the differences along the coordinates. The L_∞ (L_{inf} *or Chebychev*) gets the maximum difference of any of its coordinates.
Weighted Minkowski	$d_{L_p}(F,G) = \sqrt[p]{\sum_{i=1}^{n} w_i\,(f_i - g_i)^p}$, where w_i is the weighting vector $w_i = (w_1, w_2, ..., w_n)$	Used when there are different influences between the features that affect the similarity comparison.
Jeffrey Divergence	$d_J(F,G) = \sum_{i=1}^{n}\left(f_i \log \frac{f_i}{m_i} + g_i \log \frac{g_i}{m_i}\right)$, where $m_i = \frac{f_i+g_i}{2}$	It is symmetric and presents a better numerical behavior. Also, it is stable and robust with regard to noise and the size of histogram bins
Statistic Value χ^2	$d_{\chi^2}(F,G) = \sum_{i=1}^{n} \frac{(f_i - m_i)^2}{m_i}$, where $m_i = \frac{f_i+g_i}{2}$	It emphasizes the elevated discrepancies between two feature vectors and measures how improbable the distribution is.
Canberra	$d_C(F,G) = \sum_{i=1}^{n} \frac{\lvert f_i - g_i \rvert}{\lvert f_i \rvert + \lvert g_i \rvert}$	It is a comparative Manhattan distance, since the absolute difference in the feature values is divided by their absolute sum.

presented experiments. The color-based features were obtained from traditional histograms and Metric Histograms [Traina et al. 2002a], which are compared using the MHD distance that computes the overlapping area between two Metric Histograms acquired from two images. The texture-based extractor used carries the Haralick descriptors [Haralick et al. 1973] obtained from co-occurrence matrix. Regarding the shape-based extractors we employed Zernike moments [Khotanzad and Hong 1990], and also an improved EM/MPM algorithm proposed in [Balan et al. 2005] that segments the images using a technique that combines a Markov Random Field and a Gaussian Mixture Model to obtain a texture-based segmentation.

The contribution presented in this section aims at supporting our claim that a careful choice of a distance function considerably improves the retrieval of multimedia data [Bugatti et al. 2008a]. To do so, we performed similarity queries (*k*-nearest neighbors) on several datasets, using different distance functions, and comparing the differences on the retrieval ability. We employed images as an example of complex data, since it is widely employed in multimedia systems.

Each set of feature vectors obtained was indexed using the Metric Access Method (MAM) Slim-tree [Traina et al. 2002b], to accelerate the similarity query processing. To assess the distance function ability on properly separating the images, we have generated graphs based on precision and recall (P&R) approach [Baeza-Yates and Ribeiro-Neto 1999], obtained from the results of sets of similarity queries. A rule of thumb to read these graphs is the closer the curve to the top, the better the retrieval technique is. Therefore, the best combination of features and the distance function is achieved by the P&R curve nearest the top.

The precision and recall (P&R) graphs of Figure 1 (a) and (b) correspond, respectively, to the experiments performed on the image dataset represented by the texture-based extractor and by the shape-based extractor using the EM/MPM algorithm, considering a dataset composed of 704 images of magnetic resonance (MR) and angiogram exams, comparing all the distance functions mentioned in Table 1. The dataset was divided in 8 classes according to the region of body examined and the type of specified section: angiograms, axial pelvis, axial head, axial abdomen, coronal head, coronal abdomen, sagittal head and sagittal spine. The images were represented by 8 bits, resulting in 256 gray-levels and 256 x 256 pixels. Since the dataset was composed of 704 MR images, we processed 704 queries using each image as a query center. The average values obtained from the P&R calculation was used to generate the graphs in Figure 1.

Due to space limitations, in the present paper there are only results obtained from one dataset, and two feature extractors.

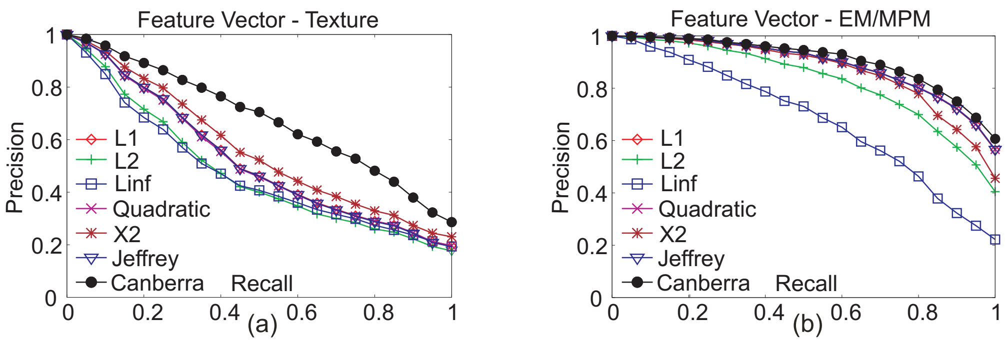

Figure 1. Precision and Recall graphs, illustrating the retrieval ability for: (a) texture and (b) shape using the EM/MPM, considering several distance functions.

The graphs in Figure 1 (a) shows that the Canberra distance presents a considerable gain in precision compared with the others, approximately 80% at a 40% of recall. The next one, the χ^2 distance, is followed by Jeffrey Divergence and L_1 to a 40% recall level, respectively resulted in a precision up to 60% and 55%. The commonly used Euclidean (L_2) and Chebychev (L_∞) presented the poorest results. The difference in precision reaches values of 92% when Canberra and Chebychev are compared. This value would make a huge difference in the response set returned to the users.

From the graphs of Figure 1 (b) we can see again that the Canberra distance presented the higher values of precision, up to 95% to a recall level of 60%. The L_2 and L_∞ gave worse results compared to the other distances. It is important to note that the difference in precision when Canberra and Chebychev are compared, reaches values of aproximately 36% at a 55% of recall. Thus, these results testify that a careful choice of a distance function improves to a great extent the precision of similarity queries (i.e the quality of retrieving complex data).

2.2. Content-based Image Retrieval by Continuous Feature Selection

CBIR systems are becoming widely used in many areas of knowledge. For instance, in medicine its main purpose is to retrieve "cases" or images similar to a given one. Analyzing past similar cases and their reports can improve the radiologist's confidence on elaborating a new image report, besides making the diagnosing process faster. Also such systems can be successfully employed in medicine teaching.

Currently, image mining has been focused by many researchers in data mining and information retrieval fields and has achieved impressive results. A major challenge of the image mining field is to effectively relate low-level features (automatically extracted from image pixels) to high-level semantics based on the human perception. Association rules have been successful applied to other research areas (e.g. business, among others) and can reveal interesting patterns relating low-level and high-level image data.

In this work, association rules were employed to support CBIR systems [Bugatti et al. 2008b]. We proposed to use association rules to weight features according to their significance, promoting continuous feature selection of the feature vectors employed to represent the images. Feature selection can significantly improve the precision of content-based queries in image databases by removing noisy features or by bursting the most relevant ones.

Feature selection techniques can employ binary or continuous approaches. The binary approach assigns binary weights to each feature, while continuous feature selection techniques assign continuous weights to each feature, allowing the most important features to have the highest weight to compute the similarity between two images. The continuous feature selection approach can improve the precision of content-based queries in a great extent.

Statistical association rules find patterns relating low-level image features to high-level knowledge from the images, and use the patterns mined to determine the weight of the features. Therefore, the feature weighting through the statistical association rules also reduces the semantic gap between low-level features and the high-level user interpretation of images, improving the precision of the content-based queries. Moreover, the proposed method performs dimensionality reduction of image features avoiding the "dimensionality curse" problem. Experiments show that the proposed method improves the precision of the query results up to 38%.

Our proposed method employs the StARMiner algorithm [Ribeiro et al. 2005] to mine statistical association rules from features of a training dataset. An important question to answer is how to use statistical association rules to weight the image features. Suppose that the images were classified in m high-level classes $X = \{x_1, x_2,..., x_m\}$. For each feature f_i, StARMiner aims at finding rules of the form $x_j \rightarrow f_i$. That is, StARMiner relates each feature f_i to each class x_j. If a rule $x_j \rightarrow f_i$ is found, it means that the feature f_i well discriminates the images from class x_j. Therefore, the most discriminative features f_i are those that generate rules $x_j \rightarrow f_i$, for every $x_j \in X$, meaning that they discriminate well all image classes. In the same way, the least discriminant features are those that do not generate any rule, meaning that they have a uniform behavior among all classes. Thus, to weight a feature f_i, the proposed method uses the number of mined rules where f_i appears. Equation 1, obtained empirically, shows the weighting assigned to each feature f_i:

$$w_i = 10 \times r_i + q \tag{1}$$

where, r_i is the number of mined rules where feature f_i appears; q is a constant that receives the values $q = 0$ or $q = 1$. The use of $q = 0$ means that it is desirable to remove the features that do not generate any rule. When $q = 1$ is used, it means that all the features are kept and weighted by relevance. Therefore, when $q = 0$, an implicit process of dimensionality reduction is performed over the feature vector, and when $q = 1$, all features are weighted according to their relevance.

The proposed method is divided in two phases: training and test. The training phase is composed of three steps: (1) feature selection, (2) association rule mining, and (3) continuous feature selection. The test phase employs the weights found in the last step of the training phase to perform similarity searches.

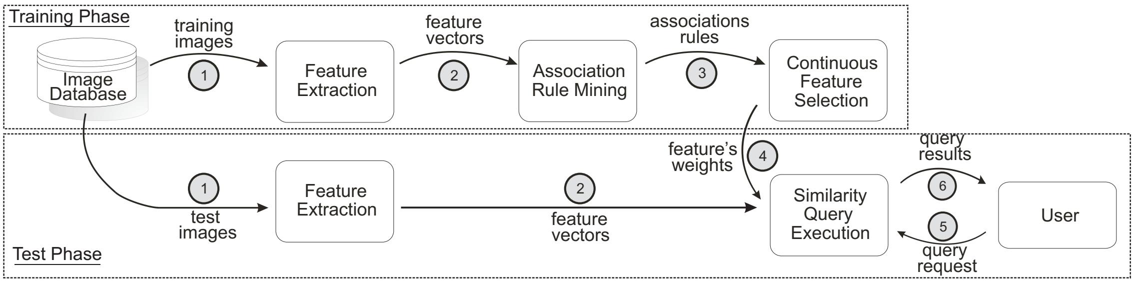

Figure 2. Pipeline of the proposed method.

Due to space limitations, in the present paper we present only results obtained from one representative dataset, the MRI dataset (see section 2.1) and using only one type of feature vector based on texture obtained from Haralick descriptors (see section 2.1), that comprises a feature vector of 140 features. In the experiments, the dataset was divided in two sets: the training set that is composed of 176 images (25% of the MRI dataset), and the test set that is composed of 528 images (75% of the MRI dataset).

The P&R graphs of Figure 3 correspond to the experiments performed on the image dataset represented by the texture-based extractor. In Figure 3 the graphs (a), (b) and (c) show the results when using L_1, L_2 and L_∞ distance functions respectively, comparing our weighting approach, with non-weighting ones. The P&R curves in the graphs of Figure 3 were built by executing similarity queries employing: (1) non-weighting the features; (2) the StARMiner feature selection; (3) the proposed method, using $q = 0$ (removing irrelevant features) and (4) the proposed method, using $q = 1$ (weighting features by relevance). In the proposed method, the use of $q = 0$ leads to dimensionality reduction of the feature vector, removing redundant features of it. For the dataset represented by the texture-based features, the use of $q = 0$ leaded to a reduction of 20% on the feature vector size (i.e the feature vectors remain with 112 features).

Analyzing the graphs of Figure 3(a) we observe that the proposed technique clearly improves the precision of similarity queries. The best precision was obtained, considering all the features and weighting them ($q = 1$). The weighted features presented a considerable gain of 30% over the performance of non-weighted features, for the recall level of 40%. When using $q = 0$, a gain of about 20% is achieved for the same recall level. It is important to note that, the use of $q = 0$ also promoted a reduction of 20% on the feature vector size, which diminishes the memory and processing cost. When analyzing the graphs of Figure 3(b), it is possible to note that the proposed technique presented a notable gain in precision. We can observe that the dimensionality reduction achieved using the StARMiner algorithm (see Figure 3(b) curve L_2 StARMiner) decays the precision values. However, when the proposed technique is applied to the feature vectors the precision increases, reaching a gain up to 20% using the $q = 0$ for a recall level of 35%, and a gain of up to 38% when $q = 1$.

These results testify that the proposed technique improves the precision of simi-

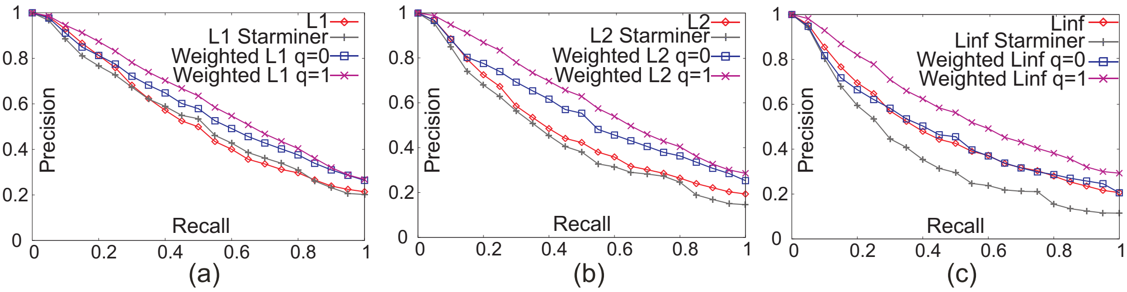

Figure 3. P & R graphs using (a)L_1, (b)L_2, and (c)L_∞ obtained over the dataset represented by texture-based features, employing: non-weighting the features; the StARMiner feature selection; the proposed method, using $q = 0$ (removing irrelevant attributes) and the proposed method, using $q = 1$.

larity queries, even when it reduces the dimensionality of feature vectors. Analyzing the graphs of Figure 3(c) we observe that StARMiner generated the worst precision even in comparison with the original features obtained by the texture extractor. When using our technique with $q = 0$ (i.e. by weighting the features selected by StARMiner) the precision increases and ties with the precision obtained by the original features. It is important to highlight that although the precisions tie, our technique also accomplished the dimensionality reduction of 20% on the feature vector size, thus demanding less space and making the processing faster. Applying our technique with $q = 1$ gave a considerable gain in precision, up to 67% for a recall level of 30% in comparison with the StARMiner algorithm, and a gain in precision, up to 20% in comparison with the precision obtained by our technique using $q = 0$ and also the original features.

3. Conclusions

The first key contribution of this MSc. dissertation demonstrated the importance of choosing the best suited distance to a kind of feature, and it was proved that having an excellent feature extraction algorithm is not a warranty of good image discrimination. The Canberra distance, a not broadly used distance and rarely employed in multimedia data retrieval, presented very promising results with regard to precision for the cases analyzed. Moreover, its computational cost is as inexpensive as the widely used distance function L_1.

The second contribution of this dissertation, was the proposal of a new supervised method of continuous feature selection. The proposed method employs statistical association rules to reduce the semantic gap inherent in the CBIR systems. Our approach can also be employed to perform dimensionality reduction, minimizing the "dimensionality curse" problem. The experiments performed show that the proposed method improves the precision of the query results up to 38%, always outperforming the precision obtained by the original features, while decreasing the memory and processing costs. These results show that statistical association rules can be successfully employed to perform continuous feature selection in image databases, weighting the features in similarity query executions.

4. Future Directions

Regarding the assessment of the best integration between distance functions and image features, we believe it opens new ways to reach the users' expectation and to enhance the acceptance and the precision of CBIR systems, since after we achieve this best combination it is also possible to branch off several other applications. For instance, it opens the door to other analysis works, as well as the integration of "perceptual parameters" to the two main elements of similarity queries (i.e. distance functions and features), resulting in more robust searching operators. Another point to be highlighted is that we can use this best integration to propose new techniques to improve the semantics of similarity queries performed by the users.

Another very promising extension of this work, regarding the proposal of the new supervised method of continuous feature selection, is that it is not confined to a specific area or weighting formula. Thus, it can be straightforwardly extended not only to other types of multimedia data and weighting techniques, but also to other distance functions.

References

Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Balan, A. G. R., Traina, A. J. M., and Marques, P. M. d. A. (2005). Fractal analysis of image textures for indexing and retrieval by content. In *18th IEEE CBMS*, pages 56–63, Dublin, Ireland.

Bugatti, P. H. and Traina, A. J. M. (2008). *Analysis of the influence of distance functions to answer similarity queries in content-based image retrieval*. Master's dissertation, University of São Paulo, Computer Science Department, São Carlos, Brazil.

Bugatti, P. H., Traina, A. J. M., and Traina Jr., C. (2008a). Assessing the best integration between distance-function and image-feature to answer similarity queries. In *Proceedings of the 23rd ACM SAC*, pages 1225–1230, New York, NY, USA. ACM.

Bugatti, P. H., Ribeiro, M. X., Traina, A. J. M., and Traina Jr., C. (2008b). Content-based retrieval of medical images by continuous feature selection. In *Proceedings of the 21st IEEE CBMS*, pages 272–277, Washington, DC, USA. IEEE Computer Society.

Deserno, T. M., Antani, S., and Long, L. R. (2007). Gaps in content-based image retrieval. *Medical Imaging 2007: PACS and Imaging Informatics*, volume 6516 of *SPIE*.

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE TSMC*, 3:610–621.

Khotanzad, A. and Hong, Y. H. (1990). Invariant image recognition by zernike moments. *IEEE TPAMI*, 12(5):489–497.

Malcok, M., Aslandogan, Y., and Yesildirek, A. (2006). Fractal dimension and similarity search in high-dimensional spatial databases. In *IEEE IRI 2006*, pages 380–384, Waikoloa, Hawaii, USA.

Ribeiro, M. X., Balan, A. G. R., Felipe, J. C., Traina, A. J. M., and Traina Jr., C. (2005). Mining statistical association rules to select the most relevant medical image features. In *IEEE MCD 2005*, pages 91–98, Houston, USA.

Traina, A. J. M., Traina Jr., C., Bueno, J. M., and Marques, P. M. d. A. (2002a). The metric histogram: A new and efficient approach for content-based image retrieval. In *Sixth IFIP VDB*, volume 216 of *IFIP Conference Proceedings*, pages 297–311, Brisbane, Australia.

Traina Jr., C.,Traina, A. J. M., Faloutsos, C., and Seeger, B. (2002b). Fast indexing and visualization of metric datasets using slim-trees. *IEEE TKDE*, 14(2):244–260.