

Using Efficient Visual Exploration Techniques to Evaluate Features for Content-based Image Retrieval

José Fernando Rodrigues Jr., César A. B. Castañón, Caetano Traina Jr., Agma J. M. Traina
Computer Science Department
University of Sao Paulo at Sao Carlos - Brazil
Avenida Trabalhador Sãocarlense, 400
13.566-590 São Carlos, SP - Brazil
e-mail: [junio, cesar, caetano, agma]@icmc.usp.br

Abstract

This paper presents a novel visual approach to evaluate, in a fast and effective way, the development of new image feature extraction techniques concerning content-based image retrieval. This new approach takes advantage of an interactive 3-dimensional visualization fed by the image features obtained through a given extraction technique under analysis. Using controlled test image datasets, the researcher can literally “see” the discrimination power of the image features. This new approach gives a very good insight of the behavior of a given image feature extractor algorithm, which are confirmed by the well-known precision and recall measurements. We applied the visual approach proposed onto a wavelet-based image retrieval system, which is supporting the development of a Picture Archiving and Communication System that allows to retrieve images by content.

1. Introduction

The ability of automatic comparison between images has been sought by the image and multimedia indexing and retrieval researching community, aiming to improve the applicability of systems that collect this kind of information. This has driven forward the development of content-based image retrieval techniques (CBIR)[1], as this is the basis for such systems. For being able to apply CBIR techniques, the first requirement is to represent the main properties, or features, of the images [2] [3]. The features are succinct representations of images that are used to perform their comparison, indexing and retrieval. Therefore, it is important that such representations really capture the “essence” of the images. However, the process of evaluating the usefulness of new image feature extraction techniques is burdensome, as it is necessary to classify the image dataset, to apply many similarity queries over the dataset and to ask

a specialist on the domain to validate the answer given by the image retrieval technique supported by the features extracted [4]. The answers given by the human specialist are compared to the results obtained by the automatic approach based on the image features, *precision* and *recall* graphs are drawn and, at this time, it is possible to compare the efficacy of the feature extractor [5].

In this paper we present a novel visual approach to evaluate the discrimination power of new image feature extraction techniques, or *extractors* for short. This approach presents the distribution of the feature vectors extracted from image databases in a three-dimensional visual space, allowing the user to manipulate it, i.e., rotating, zooming and panning the dataset in order to “see” the separation of the objects, to identify clusters and outliers, and to test the efficacy of the extractor before to spend effort evaluating it with human specialists and statistical methodologies.

In order to demonstrate the applicability of our idea, we present the results of the proposed visual approach over a wavelet-based image retrieval system, which is supporting the development of a Picture Archiving and Communication System (PACS) for medical environment.

The structure of this paper is as follows. In section 2, we discuss the approaches used for feature extraction and their applicability to CBIR systems. Section 3 discusses the role of information visualization for data analysis and the approach employed in this work. Section 4 gives the proposed idea and section 5 shows the experiments performed in order to evaluate the proposed method. Finally, section 6 presents the conclusion of this paper.

2. Background

With the increasing volume of image data generated by the medical image devices, such as magnetic resonance imaging, computerized tomography, ultrasound among others, the medicine is one of the scientific fields in more

need of effective methods for storing and retrieving images following automatic approaches. CBIR techniques have pursued this goal with relative success, but they depend strongly on the development of suitable image processing algorithms for feature extraction of images, as well as efficient access methods to index and retrieve images allowing to answer similarity queries.

2.1. Similarity Operations

For medical diagnosis based on images, it is important to retrieve cases that are similar to a given image under analysis. This allows comparisons and verifications with other diagnosis, allowing the physician to not only validate the diagnosis but also to reapply successful treatments while avoiding less effective ones.

In this paper, we considered two classes of similarity queries over images: range queries and nearest-neighbor queries, which are defined as follows.

Definition 1 (Nearest-Neighbor query): Given a query object represented by its features s_q , and the set of images S , the nearest neighbor is the unitary subset of S such that $NNquery(s_q) = \{s_n \in S \mid \forall s_i \in S: d(s_n, s_q) \leq d(s_i, s_q)\}$.

An example of a k -nearest neighbor query with $k=5$ in an image database domain is: "find the 5 nearest images to image A".

Definition 2 (Range query): Given a query object represented by its features s_q , the set of images S , and a maximum search distance r_q , the answer is the subset of S such that $Rquery(s_q, r_q) = \{s_i \mid s_i \in S: d(s_i, s_q) \leq r_q\}$.

An example of a range query is: "find all the images that are within 10 units of distance from image A", or "find the images in the database that are at least 80% similar to the given image A", where image A is the query object.

A CBIR system should allow to perform these two types of similarity queries. For that, it is necessary only two data:

- the features extracted from the images, and
- a dissimilarity (distance) function stated to compare the images, through their features.

2.2. Feature Extraction Techniques

The extraction of features is a crucial part for any image retrieval system, it provides a succinct representation (given by the feature vectors) of the images to be used in their comparison, which represent the main properties of a given image regarding a specific criteria. The most common features are based on color, shape and texture, which are employed to compare images, resulting in a score of similarity that embodies the visual differences between two determined items [1].

Now we discuss some feature extraction techniques based on the wavelet transforms [6] [7]. Further on this text, these techniques will be evaluated through our visualization approach in order to demonstrate the usefulness of the proposed method.

The intent of the feature extraction techniques is to capture the relevant characteristics of the images, reducing their dimensionality while retaining as many salient features as possible. In the present work, we chose a set of four extractors to be submitted to our approach, allowing their comparative analysis. The first chosen extractor was based on the Gabor wavelet functions [8], which capture texture measurements that represent images through their rugosity. The other three extractors were derived from the Daubechies wavelets. This wavelet function, supported by the calculus of the mean, the entropy and the energy, is used to embody image color distribution by three different approaches, as will be detailed further on.

2.3. Wavelet Transforms

It is interesting to compare the wavelet transform to the Fourier transform. While the Fourier transform analyzes a signal according to the frequency, the wavelet transform analyzes it according to the scale. Thus, the wavelets can remove statistical redundancy among pixels, providing a more compact representation of the image information. Image indexing generated over the wavelet transformed domain is believed to be more optimal than those designed over the spatial domain. This is due to the fact that the transformed coefficients have better defined distributions than image pixels [9]. Besides, the wavelets have a multi-resolution property that facilitates to extract the image features from transformed coefficients.

The central element of a multi-resolution analysis is a function $\phi(t)$, called the *scaling* function, whose role is to represent a signal at different scales. The translations of the scaling function constitute the "building blocks" of the representation of a signal at a given scale. The scale can be increased by dilating (stretching) the scaling function or decreased by contracting it.

The scaling function $\phi(t)$ acts as a sampling function (a basis), in the sense that the inner product of $\phi(t)$ with a signal represents a sort of average value of the signal over the support (extent) of $\phi(t)$. A recursive application of this process generates new nested spaces V^j , that is, $\dots V^{-2} \subset V^{-1} \subset V^0 \subset V^1 \subset \dots$, which are the basis of the multiresolution analysis.

By definition, a signal in V^{-1} can be expressed as a superposition of translations of the function $\phi^1(t)$, but because the space V^0 is included in V^{-1} , any function in V^0 can also be expanded in terms of the translations of $\phi^1(t)$. In particular, this is true for the scaling function itself.

Consequently, there must exist a sequence of numbers $\{h\}=\{h_0, h_1, \dots\}$ such that the following relationship is satisfied :

$$\phi^0(t) = \sum_n h_n \phi^{-1}\left(t - \frac{n}{2}\right) \quad (1)$$

Equation 1 is very important and is known as the scaling equation, describing how the scaling function can be generated by superposing compressed copies of itself. Now it is possible to define a new space W^j as the *orthogonal complement* of V^j in V^{j+1} . In other words, W^j is the space of all functions in V^j that are orthogonal to all functions in V^j under the chosen inner product. The relationship to wavelets lies in the fact that the spaces W^m are spanned by dilation and translation of a function $\psi(t)$, thus, such collection of basis functions are called *wavelets*. As in the case with the scaling function, since the wavelet $\psi(t)$ belongs to V^{-1} , it can be expressed as a linear combination of $\phi(t)$ at scale $m = -1$, which may be written as:

$$\psi(t) = \sum_n g_n \phi^{-1}(t - n) \quad (2)$$

where the sequence $\{g\}$ is called *wavelet sequence*. For instance, for Daubechies4 multiresolution analysis, the sequences $\{h\}$ and $\{g\}$ are defined as

$$h = \left\{ \frac{(\sqrt{3}+1)}{4\sqrt{2}}, \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}}, \frac{(\sqrt{3}-1)}{4\sqrt{2}}, -\frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}} \right\} \quad (3)$$

$$g = \left\{ -\frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}}, -\frac{(\sqrt{3}-1)}{4\sqrt{2}}, \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}}, \frac{(\sqrt{3}+1)}{4\sqrt{2}} \right\}$$

In the literature, h and g are known as the low and high frequency filters respectively.

Here, the Daubechies wavelets are used to extract features based on the color distribution over the wavelets sub-spaces, applying traditional statistical measurements. For example, given the image $I(x,y)$, with $M \times N$ pixels with depth P and its histogram $H(P)$, we have:

$$Mean = \frac{1}{M \cdot N} \sum_{i=1}^N \sum_{j=1}^M I(i, j)$$

$$Entropy = - \sum_{i=1}^P H(I) \cdot \log_2(i) \quad (4)$$

$$Energy = \sum_{i=1}^P [H(i)]^2$$

Therefore, one can extract features for evaluating the average brightness of the image (given by the mean), the

smoothness of the image (given by the entropy), and the uniformity (given by energy).

2.4. Gabor Functions and Wavelets

The use of Gabor wavelets for extracting textural features has increased in importance, because Gabor wavelets maps the modeling of simple cells in the visual cortex [10]. Gabor wavelets filters can be considered as orientation and scale tunable edge and line (bar) detectors, and the statistics of these micro-features in a given region are often used to characterize the underlying texture information. Although the Gabor functions form a complete, but non orthogonal basis set (there is redundant information in the filtered images). In [8] is proposed a technique to reduce this redundancy, which is used in this work.

A two dimensional Gabor function $g(x,y)$ and its Fourier transform $G(u,v)$ can be written as [8]:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W x \right] \quad (5)$$

$$G(u, v) = \exp \left\{ -\frac{1}{2} \left[\frac{(u - W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\} \quad (6)$$

where σ_x and σ_y are the standard deviations of $g(x,y)$ along the x and y axes, respectively. Here $\sigma_u = 1/(2\pi\sigma_x)$ and $\sigma_v = 1/(2\pi\sigma_y)$ are the standard deviations of $G(u,v)$ along the u and v axes, respectively. The Gabor function is well-known for its optimal time-frequency localization.

Gabor wavelets are generated by scaling and rotating the Gabor functions :

$$g_{mn}(x, y) = a^{-m} g(x', y'), \quad a > 1, \quad m = 1:M, \quad n = 1:N. \quad (7)$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = a^{-m} \begin{bmatrix} \cos\theta_n & \sin\theta_n \\ -\sin\theta_n & \cos\theta_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad \theta_n = n\pi / N. \quad (8)$$

where, a is the scaling parameter. Gabor wavelets can be understood as a set of Gabor functions with different frequency centers and orientations. The factor of a^{-m} yields a logarithmic frequency sampling. The size or the bandwidth of the Gabor wavelets are also controlled by a^{-m} in order to keep the energy of the Gabor wavelets constant. The orientation of the Gabor wavelets is controlled by θ . Since Gabor wavelets are symmetric, we need only to specify the value of θ to realize an evenly sampled space in $[0, \pi]$. In this way, the concept of the localization of the Gabor wavelets has been extended to time, frequency and orientation.

Given an image $I(x,y)$, its Gabor wavelet transform is defined as

$$W_{mn}(x,y) = \int I(x_1,y_1)g_{mn}^*(x-x_1,y-y_1)dx_1dy_1 \quad (9)$$

where, * indicates the complex conjugate. It is assumed that the local texture regions are spatially homogeneous, and the mean μ_{mn} and the standard deviation σ_{mn} of the magnitude of the transform coefficients are used to represent the region for classification and retrieval purposes:

$$\begin{aligned} \mu_{mn} &= \iint |W_{mn}(x,y)| dx, dy \quad \text{and} \\ \sigma_{mn} &= \sqrt{\iint [|W_{mn}(x,y)| - \mu_{mn}]^2 dx, dy} \end{aligned} \quad (10)$$

A feature vector is now constructed using μ_{mn} and σ_{mn} as feature components. For example, for two scales $S=2$ and six orientations $K=6$ the resulting feature vector, of 24 values, is composed like

$$\vec{f} = [\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \dots, \mu_{15}, \sigma_{15}] \quad (11)$$

3. Visualization Techniques

Visualization has become an indispensable tool in scientific researches to understand the generated data, to model the gained knowledge and to explain and communicate the results achieved. Appropriate visualization tools often lead scientists to quickly get new insights about the data [11]. Also, visualization techniques are capable to reveal trends and patterns hidden in data, as well as clusters and outliers, functional correlations, inconsistencies, giving a deep insight of data distribution and behavior.

The information visualization techniques have been increasingly used for analysis and exploration of large multidimensional data sets. Several approaches have been presented in the literature to achieve this intent, including the ones based on geometric projection [12], hierarchical presentation [13], graph-based [14], pixel-oriented [15] and some combinations of them [16].

In our work, a three-dimensional visualization tool is used to present information via a spatial distribution of the data under analysis. This schema can be easily comprehended by the users due to the human inherent ability to perceive the data organization in this space. The three-dimensional scene, that we use, is achieved through a mapping technique, called Fastmap [17], which was originally proposed to reduce the dimensionality of a data set.

3.1. The FastMap Technique

The Fastmap technique was developed based on the *Multi-dimensional Scaling* (MDS) [18], being able to reduce the objects' dimensionality from a n -dimensional space to a k -dimensional one, where $k < n$. The only information needed by the Fastmap are the objects and a dissimilarity function, or distance function, stated by the data domain of the objects. The FastMap mathematical details, which are not covered here, can be found on [6].

The technique works trying to maintain the distances among the objects when dimension reduction is applied. Obviously it is not possible to keep the original distance among the objects, so the FastMap tries to minimize the distortion by distributing it among the original dimensions, this is done by minimizing the average difference of the distances in between the dimensions.

In our visualization tool, any multidimensional data is reduced to 3 dimensions. Therefore, data can be plotted inside a three-dimensional Euclidean space and the user is enabled to investigate them visually and interactively. Once plotted, it becomes possible to observe the similarity among the data items, that is, the visual distance among them. In the tool, data is represented by geometrical figures of the magnitude of a three-dimensional point and, to enhance visualization, it is also possible to classify the data items by providing an attribute for data discrimination. This attribute must contain values that numerically categorize the data items by determining to which class each one of them belongs to. The plotting will use this categorization to determine form and color to differentiate visual items, so that the class they represent becomes clear to the analyst.

4. The Proposed Idea

The usual procedure to evaluate the discrimination power of a new feature extractor is to compare the image retrieval results obtained by the use of an automatic extractor with the retrieval results obtained by a specialist in the image domain. In order to make such comparison more precise, plots of the well-known *precision* and *recall* [5] measurements are drawn. *Recall* indicates the proportion of relevant images in the database which have been retrieved when answering a query. *Precision*, by the other hand, is the proportion of the retrieved images that are relevant for the query. Therefore, precision gives a direct value for measuring the retrieval ability of the method. The usefulness of a retrieval method is proportional to the value obtained, that is, the occurrence of values next to zero implies poor mapping to the user requirements, and values next to one indicates high mapping to the user specifications.

Ideally, the set of objects retrieved should be equal to the set of objects relevant for a specific query. However, mainly due to limitations on working with the features extracted

from images, and due to the distance function used to compare them, this does not happen. Through the analysis of the graphs of precision and recall one can sense the behavior of the retrieval technique to bring the correct answer for any given query. However, the generation of such plots is time consuming and demands the support of a specialist. As an alternative to such work, we purpose the use of interactive visualization to reduce the human effort.

This alternative is based on two accepted facts:

- 1) the numeric features extracted from a given image can, in fact, represent its visual characteristics. The fidelity of this representation depends on the used extractor;
- 2) visualization techniques can efficiently be used to analyze data that, in pure format, is meaningless or, otherwise, too hard to figure out.

These two facts lead us to conclude that the visualization of feature vectors could reveal their behavior, i.e., their fidelity in characterizing the images whose features were extracted. Further on, we will demonstrate that the conclusions of the visual analysis correspond to the behavior of the respective extractor.

Hence, to evaluate the usefulness of an extractor, enabling the researcher to estimate its adequacy to achieve his/her proposed goals, it is necessary to generate the visualization of the feature vectors obtained by applying this extractor. Through the visualization scene, clusters would indicate similar images, while the separation between visual elements would indicate distinct ones, as well as the perception of how different, or apart, they are.

The only condition to lead this procedure is the necessity of acquiring a set of images whose similarities/dissimilarities have already been determined. The previously analyzed images state to the analyst what to expect from any new feature extractor, consequently, the usefulness of new methods can be stipulated based on the comparison of what you get and what you expect. For a specific set of data items carefully chosen and previously analyzed, any feature extractor can be validated based on a visualization procedure rigorously carried out.

The overall idea, as depicted in figure 1, goes as follows:

- get a set of images already classified by a specialist;
- extract the image features and apply the visualization tool to the feature vectors;
- interactively visualize the feature vectors generated.

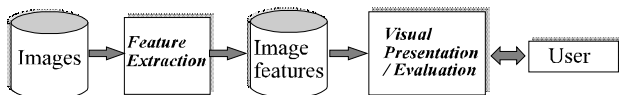


Figure 1 - Steps needed to build a visual evaluation of image feature algorithms.

Once generated the visualization scene, the analysis of the visualization goes on two main aspects: clusters and outliers. Indeed, due to the fact that the proximity in visual

elements corresponds to the proximity in feature vectors, in a straightforward correlation, the more noticeable the clusters are, the more similar are the images. On the other hand, outliers correspond to images with high dissimilarity compared to the others, presenting no correlation with them.

Thus, since the image retrieval relies on the vectors distances, just as the generated visualization, once verified the expected clusters and outliers, any feature extractor might be validated or evaluated in what matters its discrimination power. An extractor considered adequate, following the described procedure, should work well in content-based image retrieval.

4.1. Visual Analysis and Clustering

Due to the importance of cluster perception and characterization in the visual analysis, we focused on the need of an automated support for enhancing the visual investigation of the proposed method. So, to provide a way to precisely identify single and multiple clusters, we developed two add-ins on the visualization tool that are capable of numerically demonstrate these characteristics in the form of calculated metrics.

These metrics are the *variance* of a cluster, which states how condensed it is, and a boolean value, named *Mclassifier*, indicating whether or not a determined cluster defines an isolated three-dimensional subspace inside the Euclidean domain. Thus, it states whether there are elements from other classes misclassified within the present cluster. These two metrics can help to categorize the classes by measuring how good the visualized clusters are.

The *variance* of a cluster, in the Euclidean visualization context, can be thought as the spread of the elements of this class around the center of the three-dimensional subspace defined by this cluster. This subspace is the cuboid within all the elements of the cluster are contained, and it is named Minimum Bounding Box (MBB).

Once the MBB is known, what can be straightly achieved, we proceed by calculating the distance of each point of the class to the center of the box. The variance of these distances will indicate how condensed the cluster is and, consequently, the degree of similarity the feature extractor characterized the corresponding images. A well-defined cluster will have a small variance, while a widespread one will present larger values. In our tests, the variance clearly summarized the distribution of the elements of a class inside the Euclidean space. This feature is especially important in cases where two, or more, classes have a certain similarity and their elements visually overlap, hindering the clear perception of the limits of a cluster. The variance, as we verified, is also important when not efficient extractors generate diffuse visualizations where elements of various classes share the same space. Hence, as the analyst

is decisively assisted by the *variance* metric, it can be considered fundamental in the present work.

The second metric added to the tool can answer whether or not a visual cluster is completely separated from any other one. The *Mclassifier*, as we named it, is a boolean flag, so this metric will assume 0 (false) in case that the MBB of a class is invaded by elements that do not belong to this class. And, in case that the MBB defines a subspace where the points within it exclusively belong to one single determined class, the metric will assume 1 (true). The experiments showed that the proposed metric has a high potential to indicate misconception of images by the extractors. The weakness of this metric comes up with classes of images that are slightly alike, circumstance when the metric feedback cannot safely indicate the misconception in image categorization. So, while this second metric cannot be faced as a mandatory parameter for validation of feature extractors, coupled with the *variance* metric, it can make the analyst job more precise.

5. Experimental Results

Aiming to evaluate the proposed idea, we conducted experiments over image datasets that had their features generated by wavelets (Daubechies and Gabor), as described in sections 2.3 and 2.4. We first applied our method over the obtained feature vectors. Then, for confrontation, we performed nearest-neighbor queries on the image feature datasets obtaining the *precision* and *recall* measurements.

Experiments

We evaluated the proposed idea in synthetic and real datasets. Due to space limitations, we only present the results of a representative set of 500 medical images grouped in 8 classes. The 210 first images are grouped in seven classes, of 30 images each. Their medical descriptions are sagittal central human brain, abdominal, axial pelvis, axial human brain, angiographic, sagittal spine and border human brain sagittal images, respectively, as can be seen in figure 2. The eighth class, composed of 290 images is a collection of images complementary to the seven first subsets and with no inner common characteristics. This subset was added as noisy data to make the experiments more robust. Also, all the images in the subsets, except for the eighth subset, are ranked by visual similarity to one specific image pertaining its set, we called this specific image the “most relevant one” based on which the set is **ranked** to. So, the degree of equivalence between each image and the most relevant one is previously stated by a specialist, a crucial beforehand knowledge, as demanded by our methodology.

Four sets of features from the image dataset were computed: *DMean*, *DEnergy*, *DEntropy*, and *Gabor*. The first three ones are the mean, energy and entropy of the nine high-frequency orthogonal subspaces generated by the Daubechies extractor, as described in section 2.3; the fourth one was generated by the *Gabor* extractor, as explained in section 2.4.

The goal of the experiments was to compare the usefulness of the four feature extraction methods. These experiments showed the relative quality of the four extractors indicated by the *precision* and *recall* graphics, which complied with their quality expressed by the visualization analysis.

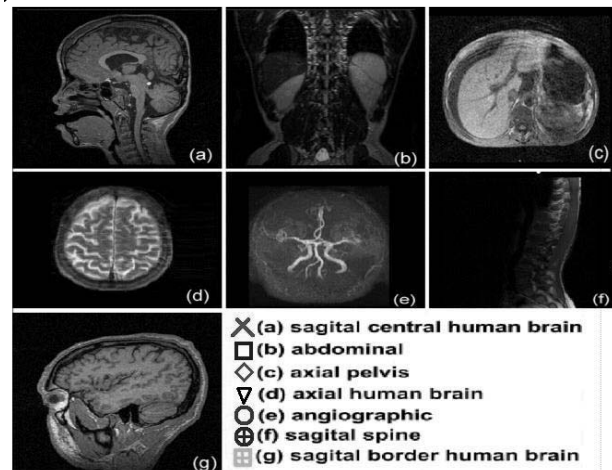


Figure 2 - The classes of images used in the experiments and their respective colors in the visualization.

The *precision* and *recall* measurements were taken for image similarity queries, as stated in section 4. The images were retrieved based on their feature vectors, using the Euclidean distance function to state their similarity. We performed sets of 5 retrieval queries for each of the 7 first subgroups of images in a 500 elements dataset, totaling 35 queries for each of the 4 extractors under analysis. In each query a random image pertaining to the 10 first images of a determined group was provided, being the query expected to retrieve the 500 elements ordered by proximity to this image.

The ideal extractor should retrieve the 30 previously classified and ranked elements of an inquired group and, in sequence, the remaining 470 images of the dataset. Here, the *precision* and *recall* methodology intends to measure how far from this idealistic scenario a given extractor is, thus a sequence of similar experiments for different extractors must show their relative efficiency, as the way we performed.

As we have evaluated four extractors, we performed 140 queries, keeping track of their results. The auditing of these results might be summarized as the recording of which of

the expected images are present in the n first elements retrieved, as stated by the technique. The average of the number of expected data items is then used to calculate the *precision* and *recall* metrics.

Results

The final *precision* and *recall* plotting is presented in figure 3. Figure 4 shows the correlation between the feature extractors visualizations, allowing to confront the results given by the *precision* and *recall* graphics. The comparison of the two techniques, through the images, goes straightly, since both of them benefits from visual perception. The more noticeable the clusters in the visualization scene are, the better the performance in the *precision* and *recall* statistics.

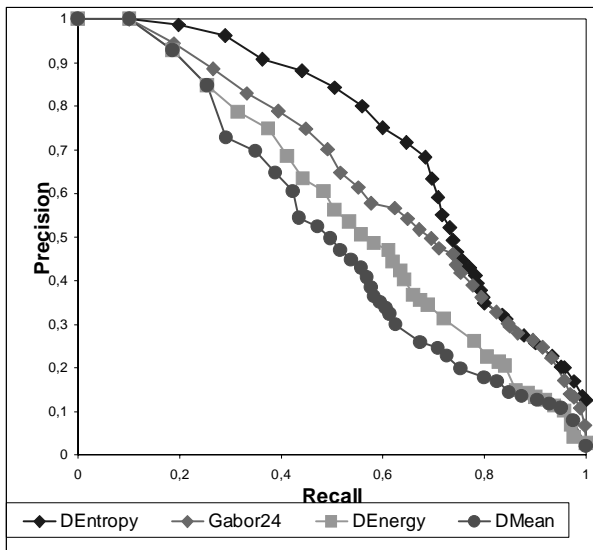


Figure 3 - The *precision* and *recall* graphs computed on the four feature extractors.

The visualization in Figure 4(a) shows that *DEntropy* gives the best separated classes, what is corroborated by figure 3, where *DEntropy* has given the best image retrieval.

The features extracted by the Gabor approach are the second in the discrimination power, as is presented in figure 4(d). Again, the *precision* and *recall* graphics (figure 3) confirms that. The other extractors, *DMean* and *DEnergy*, presented lower retrieval abilities, which is also depicted in figures 4(b) and 4(c).

To assist visual interpretation of the images in figure 4, table 1 presents the *variance* and the *Mclassifier* metrics calculated for our experiments. It can be seen that the *DEntropy* was the feature set that best categorized the images classes. Although its variance is worse than the *DMean*'s, its *Mclassifier* stated that the clusters were all

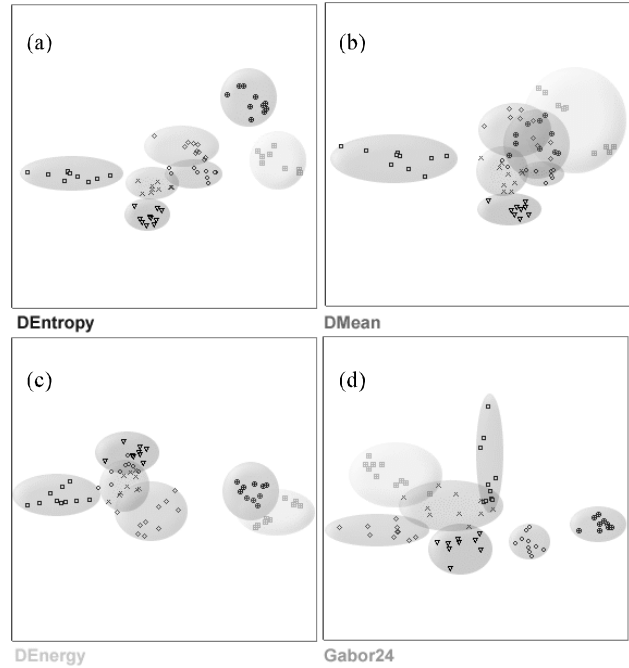


Figure 4 - The four feature extractor visualizations. Note that the colored circles delimiting the classes clusters were manually inserted for visual enhancement. The noise data is not shown.

well individualized, that is, its respective extractor can efficiently disjoin different groups of images.

Meanwhile, the opposite rules for the *DMean* features set. With a high potential to condense the clusters, its poor performance in separating them reveals that its respective feature extractor cannot tell images apart, the consequences of this fact are clearly observed in figure 3, where the misclassification of the images denoted the worst performance curve. The *Gabor* and the *DEnergy* features sets can be directly analyzed, the former demonstrated a worse performance than the *DEntropy*, and the later a slightly better result compared to the *DMean* features set. The overall analysis is in compliance with the *precision* and *recall* statistics.

	DEntropy		Gabor24		DEnergy		DMean	
	Var	Mclass	Var	Mclass	Var	Mclass	Var	Mclass
1	481.28	1	676.6	0	807.62	1	267.65	1
2	182.78	1	17.38	0	353.9	0	288.37	0
3	64.99	1	443.23	0	111.9	1	121.13	1
4	293.93	0	219.59	1	394.18	0	162.19	0
5	143.48	1	45.85	1	261.55	1	82.44	0
6	245.65	1	70.75	1	649.78	0	231.59	0
7	97.41	0	388.3	1	121.95	0	78.98	0
Sum	1509.52	5	1841.7	4	2700.88	3	1232.4	2
Avg	215.6457		263.1		385.84		176.05	

Table 1 - Variance and Mclassifier for the seven groups, calculated for each of the four feature extractors.

6. Conclusions

In this paper we presented a powerful but easy to use technique to evaluate the usefulness of image feature extractor algorithms for content-based image retrieval applications. The technique is based on the visualization of the feature vectors distribution on a three-dimensional space. The results of our experiments were very promising, being able to determine the efficiency of four different feature extractors, whose *precision* and *recall* analysis is in accordance with what was demonstrated by our visualization tool. It is important to emphasize that the features visualization is achieved instantaneously and provides a good insight of the features discrimination power, and complies with the *precision* and *recall* analysis.

We also proposed the use of the *variance* and the *Mclassifer* metrics as automatic measurements to indicate the degree of separability among clusters given by the extractor under analysis. These metrics, applied on the visualization tool, have demonstrated to be very effective to assist the visual analysis. Coupled, the metrics can clarify the perception of the clusters by denoting how condensed they are and, at the same time, express the capacity of an evaluated extractor in the task of disjointing classes of images. Although this early work was successful, the automated analysis provided herein can be widely incremented by the large range of analytical possibilities of the Euclidean space.

Acknowledgments

The research presented in this paper was supported by the Sao Paulo State Research Foundation (FAPESP) under grants 01/11287-1, 02/07318-1, and by the Brazilian National Research Council (CNPq) under grants 52.1685/98-6, 860.068/00-7 and 35.0852/94-4.

7. References

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, 2000.
- [2] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Transactions on Image Processing*, vol. 10, pp. 117-30, 2001.
- [3] A. K. Jain and A. Vailaya, "Shape-Based Retrieval: A Case Study with Trademark Image Databases," *Pattern Recognition*, vol. 21, pp. 1369-1390, 1998.
- [4] H. Yamamoto, H. Iwasa, N. Yokoya, and H. Takemura, "Content-based Similarity Retrieval of Images Based on Spatial Color Distributions," presented at 10th Intl. Conference on Image Analysis and Processing, pp. 951-956, 1999.
- [5] R. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval*. Wokingham, UK: Addison-Wesley, 1999.
- [6] J. Z. Wang, "Wavelets and Imaging Informatics: A Review of the Literature," *Journal of Biomedical Informatics*, vol. 34, pp. 129-141, 2001.
- [7] E. Albu, E. Kocalar, and A. A. Khokhar, "Scalable Color Image Indexing and Retrieval Using Vector Wavelets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, pp. 851-861, 2001.
- [8] B. S. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837-842, 1996.
- [9] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, *Wavelets for Computer Graphics - Theory and Applications*. San Francisco, CA: Morgan Kaufmann Publishers, Inc, 1996.
- [10] T. S. Lee, "Image Representation Using 2D Gabor Wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, 1996.
- [11] K.-L. Ma and F. H. Post, "Visualization Case Studies: Drawing a Roadmap for Future Visualization," *IEEE Computer Graphics and Applications*, vol. 19, pp. 48-49, 1999.
- [12] A. Inselberg and B. Dimsdale, "Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry," presented at IEEE Visualization, pp. 361-370, 1990.
- [13] Y.-H. Fua, M. O. Ward, and A. Rundensteiner, "Hierarchical Parallel Coordinates for Exploration of Large Datasets," *Proc. IEEE Visualization'99*, 1999.
- [14] R. A. Becker, S. G. Eick, and G. J. Wills, "Visualizing Network Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, pp. 16-28, 1995.
- [15] D. A. Keim, "Designing Pixel-Oriented Visualization Techniques: Theory and Applications," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, pp. 59-78, 2000.
- [16] D. A. Keim, "Information Visualization and Visual Data Mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, pp. 1-8, 2002.
- [17] C. Faloutsos and K.-I. D. Lin, "FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets," presented at ACM Intl. Conference on Data Management (SIGMOD), San Jose, CA, pp. 163-174, 1995.
- [18] W. S. Torgenson, "Multidimensional Scaling: I. Theory and Methods," vol. 17: *Psychometrika*, 1952, pp. 401-419.